

Dissertation  
submitted to the  
Combined Faculty of Natural Sciences and Mathematics  
of the Ruperto Carola University Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

Presented by  
M.Sc. Tomi Määttä

Born in: Kuusamo, Finland

Oral examination: 16.09.2020

## **Stress response of the proteome**

Referees: Dr. Athanasios Typas

Prof. Dr. Bernd Bukau

# TABLE OF CONTENTS

AKNOWLEDGEMENTS .....	vii
SUMMARY .....	viii
ZUSAMMENFASSUNG .....	ix
LIST OF FIGURES .....	xi
ABBREVIATIONS .....	xiii
INTRODUCTION .....	1
1. Protein folding .....	1
2. Protein misfolding .....	3
2.1 Changes in protein sequence .....	4
2.2 Proteotoxic stress .....	5
2.3 Heat shock .....	6
2.4 Lowered capacity of protein quality control .....	7
3. Protein aggregation .....	7
3.1 Protein conformations in aggregates .....	8
3.2 Phase separation .....	9
3.3 Spatial aggregation .....	10
4. Processing of protein aggregates .....	12
4.1 Proteasomal degradation .....	12
4.2 Autophagy .....	14
4.3 Disaggregation .....	15
4.4 Aggregate secretion .....	17
5 Protein aggregation and disease .....	17
6. Mass spectrometry-based proteomics in protein aggregation studies .....	18
6.1 Principles of mass spectrometry .....	19
6.2 Mass spectrometry-based proteomics .....	19
6.3. Quantitative mass spectrometry .....	20
6.3.1 Protein level labelling by SILAC .....	21
6.3.2 Peptide level labelling by tandem mass tags .....	22

6.3.3 Hyperplexing - combination of labelling techniques .....	23
6.4 Mass spectrometry in protein aggregation studies .....	23
AIMS AND OBJECTIVES.....	25
1. Aims .....	25
2. Objectives .....	26
2.1 Development of mass spectrometry-based platform to study protein solubility after heat shock and during recovery .....	26
2.2 Characterization and quantification of proteins prone for aggregation in heat shock .....	26
2.3 Characterization and quantification of protein disaggregation during recovery from heat shock.....	27
2.4 Analysis of heat shock-induced changes in protein synthesis.....	27
2.5 Analysis of heat shock-induced changes in thermal stability of non-aggregating proteins .....	27
MATERIALS AND METHODS .....	28
1. Cell culture .....	34
2. Preparing cells for heat treatments.....	34
3. Heat treatment .....	36
3.1 Water bath .....	36
3.2 Heat block .....	36
4. Sample collection .....	37
5. Viability assay .....	37
6. Two dimensional proteome profiling.....	38
7. Cell lysis .....	39
8. Removal of insoluble fraction .....	39
9. Protein extraction .....	40
10. Tryptic digestion .....	41
11. TMT labeling .....	41
12. Peptide desalting.....	42
13. Off-line fractionation .....	42
14. Quantitative mass spectrometry.....	42
15. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis .....	44
16. Data analysis.....	44

17. Protein features .....	52
RESULTS.....	53
1. Experimental design.....	53
1.1 Dynamic stable isotope labeling by amino acids in cell culture .....	53
1.2 Pilot experiment to measure aggregation and disaggregation .....	54
1.3 Protein aggregation in K562 cells .....	55
1.4 Aggregation in updated heat shock method.....	57
1.5 Finalized design: heat treatment, sample collection and quantitative mass spectrometry .....	58
2. Proteins aggregating in transient heat shock .....	61
2.1 Aggregators are enriched in nuclear proteins .....	63
2.2 Stress granule proteins remained soluble after heat shock .....	65
2.3 Insoluble sub-populations .....	67
2.4 Physicochemical properties of aggregators .....	70
2.5 Structural features of aggregators.....	72
2.6 Protein complexes and aggregation.....	75
3. Protein disaggregation during recovery from heat shock .....	77
3.1 Wide-spread disaggregation during recovery.....	77
3.2 Different disaggregation kinetics relate to different nuclear functions .....	80
3.3 Disaggregation is driven by aggregation propensity and amount of intrinsically disordered regions .....	82
3.4 Protein complexes and disaggregation .....	83
4. Heat shock-induced effects on protein synthesis .....	85
4.1 Reversible stall in protein synthesis after heat shock.....	85
4.2 Upregulation of protein synthesis after heat shock .....	88
5. Thermal stability changes of soluble proteins.....	90
5.1 Thermal stabilization of soluble remnants of aggregators.....	92
5.2 Thermal stability changes and stress signaling.....	94
5.3 RNA polymerase II destabilized upon heat shock.....	95
5.4 Destabilization of H1 histones.....	96
5.5 Comprehensive destabilization of quality control protein complexes .....	96
DISCUSSION.....	98
1. Nucleus is sensitive for proteotoxic stress.....	99

2. DNA-binding proteins contribute to the nuclear sensitivity .....	102
3. Aggregation and disaggregation of protein complexes.....	103
4. Protein aggregation to quality control sites in nucleus.....	104
5. Intrinsically disordered regions protect from toxic interactions and facilitate disaggregation.....	106
6. The aggregation of proteins with high molecular weight, hydrophilic character and high pI .....	108
7. Slow disaggregation in metazoans.....	110
8. Soluble remnants of aggregating proteins are protected from aggregation.....	112
REFERENCES.....	114

## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Misha Savitski, a genius scientist, whose guidance has enabled me to grow with and execute this project.

From the bottom of my heart, I would like to thank Mandy Rettel a million times for the technical and mental support that has been vital for this project.

I am forever grateful for the crucial input from Frank Stein on data analysis, experimental design and encouraging feedback.

My warmest gratitude goes to Isabelle Becher, who's unconditional help and support has been priceless. A special thank you for baby Nora Becher for delightful discussions and laughter.

I will never forget the remarkable office environment created by Sindhuja Sridharan, Nils Kurzawa and Henrik Hammarén. Thank you very much for all the support and sparkling discussions related to research projects and beyond.

I am deeply grateful for André Mateus for his efforts and support in countless matters.

Special thanks for Per (and Sara and Linus) Haberkant for friendliness, discussions and support that have been more important to me than one might imagine.

I would like thank Dominic Helm for ensuring excellent performance on mass spectrometers and for all of his support.

I want to thank Cosimo Jann for lively discussions on every aspect of protein quality control.

I would like to show my gratitude to Jianguo Zhu for the great support, feedback and amazing discussions.

Big thanks to all (past and present) members and associates of the Savitski team for support and insightful discussions: Clément Potel, Johannes Hevler, Malay Shah, Dylan Mooijman and Vallo Varik.

I highly appreciate the input from my thesis advisory committee. Therefore, I would like to thank Nassos Typas, Bernd Bukau and Martin Beck.

I would like to thank EMBL International PhD Program for funding and Matija Grgurinovic, Carolina Sabate and Meriam Bezohra from the graduate office for their great support.

Finally, I would like to show my deepest gratitude to my family.

## SUMMARY

Neurodegenerative diseases are often associated with the formation of protein aggregates. While this association is well established, the causal link between disease progression and protein aggregation is largely unknown. Protein aggregation processes have been mainly studied using reductionist experimental systems. However, it has recently become evident that protein aggregation and handling of protein aggregates is an actively controlled process in the cellular environment. Thus, it is crucial to study protein aggregation in a cellular context.

In this project, a mass spectrometry-based proteomics workflow was established to study protein aggregation, disaggregation and synthesis of endogenous human proteins *in situ* upon transient and non-lethal heat shock. To extend the scope, changes in thermal stability of proteins that remained soluble after heat shock were analyzed by thermal proteome profiling.

It was found that transient heat shock induced the aggregation of 300 mainly nuclear proteins enriched in intrinsically disordered regions, hydrophilic amino acids, high molecular weight and high isoelectric point. During recovery, most aggregated proteins became disaggregated. The disaggregation rates were found to correlate with the amount of intrinsically disordered regions in the proteins but not with other features enriched in aggregating proteins. In addition, larger loss of solubility after heat shock was counteracted by faster disaggregation. Protein synthesis had a global reversible halt after heat shock followed by an upregulation of heat shock proteins. Thermal stability was increased for soluble remnants of aggregating proteins, suggestive of a protective mechanism that prevents complete aggregation of unstable proteins. Furthermore, heat shock induced changes in thermal stability for proteins related to stress signaling, DNA binding and quality control.



## ZUSAMMENFASSUNG

Neurodegenerative Erkrankungen gehen häufig mit der Bildung von Proteinaggregaten einher. Obwohl dieses Phänomen häufig beschrieben wurde, ist der ursächliche Zusammenhang zwischen Krankheitsprogression und Proteinaggregation weitgehend unbekannt. Proteinaggregationsprozesse wurden bisher hauptsächlich in reduktionistischen experimentellen Systemen untersucht. Kürzlich hat sich jedoch gezeigt, dass die Proteinaggregation und der Umgang mit Proteinaggregaten in der zellulären Umgebung ein aktiv kontrollierter Prozess ist. Daher ist es wichtig, die Proteinaggregation im zellulären Kontext zu untersuchen.

In diesem Projekt wurde ein auf Massenspektrometrie basierender Proteomics-Workflow eingerichtet, um die Proteinaggregation, -disaggregation und -synthese endogener menschlicher Proteine in situ bei vorübergehendem und nicht letalem Hitzeschock zu untersuchen. Zusätzlich wurden Änderungen der thermischen Stabilität von Proteinen, die nach einem Hitzeschock löslich blieben, durch thermisches Proteomprofilieren (Thermal proteome profiling) analysiert.

Die Resultate ergaben, dass ein vorübergehender Hitzeschock die Aggregation von 300 hauptsächlich nuklearen Proteinen induzierte. Diese Proteine waren angereichert an intrinsisch ungeordneten Regionen, hydrophilen Aminosäuren, hohem Molekulargewicht und hohem isoelektrischen Punkt. Nach dem Hitzeschock wurden die meisten aggregierten Proteine von der Zelle disaggregiert. Die Proteindisaggregationsraten korrelierten mit der Menge an intrinsisch ungeordneten Regionen, jedoch nicht mit anderen Merkmalen, die an aggregierenden Proteinen angereichert waren. Größere Löslichkeitsverluste nach Hitzeschock wurden zusätzlich von der Zelle durch eine schnellere Disaggregation ausgeglichen. In der Proteinsynthese konnte nach einem Hitzeschock eine globale, reversible Unterbrechung, gefolgt von einer Hochregulierung der Hitzeschockproteine erkannt werden. Die thermische Stabilität löslicher Reste

aggregierender Proteine schien erhöht, was auf einen Schutzmechanismus hindeutet, der eine vollständige Aggregation instabiler Proteine verhindert. Darüber hinaus verursachte ein Hitzeschock Änderungen der thermischen Stabilität von Proteinen im Zusammenhang mit Stresssignalisierung, DNA-Bindung und Proteinqualitätskontrolle.

## LIST OF FIGURES

- Figure 1      Reproducibility between replicates in light SILAC dataset.
- Figure 2      Reproducibility between replicates in heavy SILAC dataset.
- Figure 3      Reproducibility between replicates in light SILAC dataset. Data from dynamic SILAC experiment with heat shock.
- Figure 4      Reproducibility between replicates in heavy SILAC dataset. Data from dynamic SILAC experiment with heat shock.
- Figure 5      Reproducibility between replicates in two-dimensional thermal proteome profiling.
- Figure 6      Protein aggregation and disaggregation examined by quantitative mass spectrometry.
- Figure 7      Protein aggregation in K562 cells examined by SDS-PAGE.
- Figure 8      Protein aggregation at different heat shock temperatures examined by quantitative mass spectrometry.
- Figure 9      Experimental design to study protein solubility after heat shock and during recovery.
- Figure 10     Mass spectrometry analysis of dynamic SILAC data.
- Figure 11     Solubility change after heat shock.
- Figure 12     Functional characterization of aggregators.
- Figure 13     Localization of aggregators.
- Figure 14     Solubility of stress granule proteins.
- Figure 15     Insoluble protein sub-populations.
- Figure 16     Physicochemical features of aggregators.
- Figure 17     Structural features of aggregators.
- Figure 18     Aggregators within protein complexes.
- Figure 19     Solubility of aggregators within protein complexes.

- Figure 20 Disaggregation of aggregated proteins during recovery from heat shock.
- Figure 21 Quantification of disaggregation.
- Figure 22 Total protein levels from pre-existing fraction stays constant during recovery from heat shock.
- Figure 23 Disaggregation kinetics.
- Figure 24 Molecular features of aggregators and disaggregation kinetics.
- Figure 25 Disaggregation of protein complex members.
- Figure 26 Protein synthesis after heat shock.
- Figure 27 Total protein levels of newly synthesized proteins during recovery from heat shock.
- Figure 28 Newly synthesized proteins aggregating upon heat shock.
- Figure 29 Upregulation of protein synthesis after heat shock.
- Figure 30 Schematic presentation for two-dimensional thermal proteome profiling.
- Figure 31 Heat shock-induced changes in thermal stability of soluble proteins.
- Figure 32 Soluble remnants of aggregators stabilized upon heat shock.
- Figure 33 Heat shock-induced stability changes in stress-activated MAPK pathway.
- Figure 34 Destabilization of proteins related to DNA binding.
- Figure 35 Destabilization of protein quality control complexes upon heat shock.

## ABBREVIATIONS

2D-TPP	Two-dimensional proteome profiling
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CCT	Chaperonin containing TCP-1
CO <sub>2</sub>	Carbon dioxide
DNA	Deoxiribonucleic acid
dSILAC	Dynamic stable isotope labelling of amino acids in cell culture
ER	Endoplasmic reticulum
FBS	Fetal bovine serum
FDR	False discovery rate
GO	Gene ontology
INQ	Intranuclear quality control compartment
IPOD	Insoluble protein deposit
JNK	c-Jun N-terminal kinase
LC	Liquid-chromatography
MAPK	Mitogen-activated protein kinase
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
PBS	Phosphate-buffered saline
Pol-II	RNA-polymerase-II
RNA	Ribonucleic acid
RT	Room temperature
SDS	Sodium dodecyl sulfate
SILAC	Stable isotope labelling of amino acids in cell culture
TMT	Tandem mass tags
tRNA	Transfer ribonucleic acid

# INTRODUCTION

## 1. Protein folding

To function properly, proteins need to acquire a three-dimensional fold state that is adequate to drive their function. For example, active site of an enzyme can be composed from protein regions that are far away from each other in the amino acid chain; for the enzyme to function, these regions need to be brought together in an organized manner. On the other hand, some proteins do not have a specified three-dimensional structure (Hartl *et al*, 2011). As a protein is synthesized on ribosome, the unfolded amino acid chain has higher energy than the native fold (Dinner *et al*, 2000). Therefore, it would spontaneously (in thermodynamic sense) proceed towards the native fold. The main driver in protein folding is the hydrophobic effect - burial of hydrophobic regions inside the protein to minimize contact with the bulk water solvent (Hartl & Hayer-Hartl, 2009). Buried hydrophobic regions - as well as exposed hydrophilic regions - lower the energy of the protein system.

Small proteins can reach their native fold on their own while especially larger proteins with more complex domain structure require chaperones or other folding catalysts for folding (Anfinsen, 1973, Vabulas *et al*, 2010). As larger proteins contain longer amino acid chains, the number of possible conformations they could adopt is much higher. Therefore, larger and more complex proteins are prone to have non-functional folding intermediates that might have lower energies (Dinner *et al*, 2000). Proteins can get trapped to these so called local energy minima and would have lower probability of obtaining the native fold on their own. The native fold might be of lower energy but an energy barrier in the local minima prevents a protein from continuing its folding towards the native state.

The energy barrier that a trapped folding intermediate would need to overcome can be lowered by interactions with molecular chaperones (Balchin *et al*, 2016, Tyedmers *et al*, 2010). In addition, chaperons can prevent the protein from adapting fold states corresponding to the low energy folding traps. (Balchin *et al*, 2016, Tyedmers *et al*, 2010). For example, the Hsp60 chaperonin (a special name give for this type of chaperone) is thought to close the substrate protein to a closed chamber where the walls of the chamber are changed from mainly hydrophobic to mainly hydrophilic during the chaperonin cycle (Mayer, 2010). This puts the protein in to an environment that prevents hydrophobic contact with other proteins (Hartl *et al*, 2011). The hydrophobic contacts with other proteins could stabilize the folding intermediate and, therefore, interfere with the folding.

The meta-stability of proteins discussed above means that on their own proteins would be in danger of not acquiring their native fold or obtaining a non-functional (or even toxic) conformations. Cells have specialized proteins that maintain the healthy state of the proteome. This protein quality control system comprises approximately 800 proteins (Hartl *et al*, 2011). The number of proteins involved in it highlight how important and versatile the system is. The most acknowledged group is molecular chaperones that assist in many essential processes related to adequate protein function, such as folding, transport and degradation (Vabulas *et al*, 2010). The major chaperone classes are Hsp100, Hsp90, Hsp70, Hsp60, Hsp40 and small heat shock proteins (Hartl *et al*, 2011); some of them are discussed in more detailed later. Other components in the protein quality control involve protein degradation machineries, such as ubiquitin-proteasome system and autophagy (Balchin *et al*, 2016). Therefore, the quality control system can assist in all stages in the life cycle of a protein. In addition to maintaining the fold state of proteins (and degrading damaged proteins), the quality control has an important role in keeping and adjusting the protein levels - a process called protein turnover (Hinkson & Elias, 2011).

In *de novo* protein folding (i.e. folding of newly synthesized proteins), chaperone classes operate sequentially forming a chaperone pathway (Balchin *et al*, 2016). This hierarchical arrangement of chaperones can be found from bacteria to humans (Hartl *et al*, 2011). In brief, as proteins are synthesized at ribosomes, they are first encountered by different Hsp70 systems; these include trigger factor (in bacteria), and nascent chain-associated complex and ribosome-associated complex (in eukaryotes) which are located near the ribosome exit tunnel (Balchin *et al*, 2016). The Hsp70 systems at the ribosome are followed by another Hsp70 system that is not attached to ribosome, finally followed by Hsp90 or Hsp60 as the folding progress. However, since each protein has its unique native fold, the need for the assistance of chaperones varies from protein to protein; the number of proteins needing a specific chaperone system decreases along the chaperone pathway (Balchin *et al*, 2016).

Protein folding in endoplasmic reticulum (ER) and mitochondria follow similar principles as described above (Braakman & Hebert, 2013, Buchberger *et al*, 2010, Hartl *et al*, 2011, Voos *et al*, 2016). Both compartments have their own set of chaperones from the main chaperone classes (excluding small heat shock proteins) (Buchberger *et al*, 2010, Voos *et al*, 2016). However, ER is lacking Hsp60 chaperonin (Buchberger *et al*, 2010). Many post-translational modifications are made in ER, such as oxidation of two cysteine residues to form covalent disulphide bonds between them (Hartl *et al*, 2011). Therefore, ER has an additional set of folding enzymes that assist in post-translational modifications (Braakman & Hebert, 2013).

## **2. Protein misfolding**

The native fold of a protein is not stable and is prone for adapting other non-functional low energy fold states. In addition to *de novo* folding, it is relevant especially during proteotoxic stress or other environmental changes. The non-functional low energy



states can expose hydrophobic residues that in the native fold would be buried inside the protein. This could lead to increased hydrophobic contacts between the proteins and cause proteins to form larger aggregates. (Tyedmers *et al*, 2010, Vabulas *et al*, 2010)

## 2.1 Changes in protein sequence

Genetic mutations can change the primary amino acid sequence of a protein. Mutations in protein coding region can lead to a loss of function (Hartl *et al*, 2011). In terms of folding, mutations could pass their loss of function effects by changing the energy state of non-functional local energy minima or the native fold. Changed amino acid sequence has the potential to create a new non-functional low energy fold state, lower the energy of an existing one or increase the energy barrier needed to get out from it. These changes would make a protein more prone for being trapped in the local energy minima. Additionally, the energy of the native state might increase, making it less stable. In this case, the protein would be less likely to adapt the native fold and more likely to adapt some other (non-functional) low energy state.

For a mutation to have an effect on a protein folding or function it does not require a large change in the amino sequence. A change in only one residue can have drastic outcomes. For example, a glutamate to valine point mutation in  $\beta$ -globulin chain changes the conformation of hemoglobin and exposes hydrophobic regions that subsequently interact with other hemoglobin molecules and severely hampers the function of hemoglobin (Gibson & Ellory, 2002). In general, genetic mutations are linked to many diseases where the pathology stems from protein misfolding (Nillegoda *et al*, 2018, Valastyan & Lindquist, 2014).

A change in the amino acid sequence can stem from other sources than a mutation in the deoxyribonucleic acid (DNA) sequence. The information transfer from DNA to

messenger ribonucleic acid (mRNA) and eventually to proteins is prone for errors. Therefore, changes in primary sequence of a protein could arise from these errors in addition to mutations in the DNA sequence. Indeed, the protein sequence can be altered without underlying genetic mutations by mistakes in protein synthesis (Drummond & Wilke, 2008). The impact on protein folding of incorporating wrongly assigned amino acids to the growing amino acid chain is utilized in research: by adding non-natural amino acids to proteins with the aim to interfere with the folding has been used to study protein aggregation. For example, azetidine-2-carboxylic acid can be used as a proline analog that causes protein misfolding and aggregation (Weids *et al*, 2016).

## **2.2 Proteotoxic stress**

The native fold of a protein can be disturbed by proteotoxic stress or other environmental changes. As genetic mutations directly impact only one protein species, proteotoxic stress potentially effects all proteins. These stress condition include, for example, heat stress, cold stress and changes in osmolarity or pH (Vabulas *et al*, 2010).

Different organisms can tolerate the stress conditions to different extent. For example, the bacterial proteome can tolerate higher temperatures than human as indicated by higher protein melting points (Leuenberger *et al*, 2017, Mateus *et al*, 2018). The physical stress conditions can shift the equilibrium from favoring the native state towards non-functional states; the shift can be caused directly by the changed environment or indirectly from post-translational modifications caused by the changing environment (Vabulas *et al*, 2010). For example, heat stress can directly unfold a protein but it can also cause a rapid change in the phosphorylation status of many proteins (Kanshin *et al*, 2015). The post-translational modification can have similar effects as mutations in the protein sequence: in both cases, the physicochemical features of amino acids in the proteins are altered.

## 2.3 Heat shock

The increased temperature in heat shock means an energy input that spreads throughout the cell and thus effects all proteins. Increase in temperature means faster movements of particles (i.e. kinetic energy). When this energy is absorbed by proteins, the conformation of the whole protein can start to change (Sang *et al*, 2016). This can be seen as reversal of the folding process (possibly uncontrollably) and can cause protein misfolding, exposure of hydrophobic regions and aggregation. Since the heat stress is wide-spread, the chaperone machinery might be overwhelmed and not able to handle all proteins that would need their assistance in avoiding or getting out from the non-native fold states. This diminished chaperone capacity can further speed up the process.

In addition to its impact on protein (mis)folding, heat shock induces a range of cellular responses. These include adaptive responses to the stress as well as damage caused by the elevated temperature. However, it can be challenging to distinguish between the two. For instance, the expression of genes is rewired on the levels of transcription (Mahat *et al*, 2016, Vihervaara *et al*, 2017), mRNA splicing (Yamamoto *et al*, 2016), protein synthesis (Muhlhofer *et al*, 2019) and post-translational modifications (Kanshin *et al*, 2015). Heat shock induces the expression of so called heat shock proteins (De Maio *et al*, 2012, Ritossa, 1962, Ritossa, 1996) that include chaperones from all major classes as well as other protein quality control components, such as members of the ubiquitin-proteasome degradation system (Kovács *et al*, 2019). This increased gene expression can be tuned to proceed while other transcription and translation is stopped. For example, the translation machinery can be adapted to selectively process mRNA of heat shock proteins (Wallace *et al*, 2015). The induced expression of chaperones not only assists in coping with the experienced heat stress, but also prepares and makes cells less vulnerable for subsequent stress - a state called acquired thermotolerance (Sanchez & Lindquist, 1990). Therefore, the cellular response to heat stress is highly organized on the level of gene expression.

While changes in gene expression are mainly observed on molecular level, the morphology of larger cellular organs might undergo changes upon heat shock. Changes can be observed on many levels ranging from whole cell to macromolecules that are visible with light microscopy. On the highest structural level, the shape of the whole cell can change upon heat shock (Schamhart *et al*, 1984). The major organelles, such as mitochondria (Funk *et al*, 1999, Welch & Suhan, 1985) and Golgi (Welch & Suhan, 1985) have been observed to change shape upon heat shock (Welch & Suhan, 1985). The smaller compartments with changed morphology upon heat stress include nuclear structures, such as chromatin (Ritossa, 1962) (an observation that eventually led to the finding of heat shock proteins) or nucleolus (Pelham, 1984). Together these observations highlight the wide-spread cellular impact a heat stress can cause.

## **2.4 Lowered capacity of protein quality control**

Chaperones assist in protein folding and other processes related to having functional proteins (Hartl & Hayer-Hartl, 2009). Therefore, decline in the chaperone capacity (or other proteins related to maintaining functional proteins) is a cause for protein misfolding. The chaperone capacity can be overwhelmed by proteotoxic stress (as discussed above in the context of heat shock); misfolded proteins accumulate that require more chaperones (Tyedmers *et al*, 2010). Eventually, the chaperone capacity is used up and unable to cope with the increasing numbers of misfolded proteins. Interestingly, beside direct proteotoxic stress, aging has been shown to decline the chaperone capacity (Hartl, 2016, Walther *et al*, 2015).

## **3. Protein aggregation**

Misfolded proteins with exposed hydrophobic regions can initiate protein accumulation to aggregates (Scior *et al*, 2016, Tyedmers *et al*, 2010). As with protein folding, the

aggregation is thought to be driven by hydrophobic effect (Hartl & Hayer-Hartl, 2009); hydrophobic contact between proteins shields the exposed hydrophobic regions from the bulk water solvent and thus lowers the energy of the protein system.

### **3.1 Protein conformations in aggregates**

Proteins can adopt quite different conformations in different kinds of aggregates. Therefore, there is no one type of protein deposit that can be found after protein aggregation. This adds complexity to the mechanisms how aggregates form, possibly develop over time and are handled by the cells. However, some general features can be found. For example, the  $\beta$ -sheet content is typically higher in proteins at the aggregates (Tyedmers *et al*, 2010). The conformations that proteins adapt in aggregates has also been grouped to amorphous aggregates and amyloid fibers (Scior *et al*, 2016), although this view has been challenged as over-simplification (Tyedmers *et al*, 2010).

As the name suggests, amyloid fibers form fiber-like aggregates that can be visualized with electron microscopy (Scior *et al*, 2016). The amyloid fibers are characterized by  $\beta$ -sheets that are perpendicular to the direction of the fiber (forming so called cross- $\beta$ -sheets) (Wang *et al*, 2010). In addition, amyloid fibers are very stable structures that often have much lower energy than the native fold of a protein (Hartl & Hayer-Hartl, 2009). The stability of a protein aggregate can be so strong that they do not dissolve in strong detergents used to normally solubilize proteins (Hosp *et al*, 2017). However, as discussed above, the distinction to amorphous aggregates and amyloid fibers might reflect two extremes of the same continuum and amorphous aggregates can also form fibers and contain cross- $\beta$ -sheets.

The conformations proteins adopt in aggregates can dependent on the perturbation that caused the misfolding and aggregation. Even the same protein species can form

amorphous aggregates or amyloid fibers (Chiti & Dobson, 2017), which can be dependent on the stress conditions that caused the aggregation (Wang *et al*, 2010). For example, when precipitated with trichloroacetic acid, N-terminal domain of the hydrogenase maturation factor HypF forms amorphous aggregates with little increase in  $\beta$ -sheet content while after heat shock it formed more stable aggregates with more increased  $\beta$ -sheet content (Wang *et al*, 2010).

Proteins can also form aggregates where they are at or near native states (Chiti & Dobson, 2017). The aggregated proteins can even perform their functions as illustrated by yeast multisynthetase complex that after aggregating upon heat shock, had no change in its ability to aminoacylate transfer RNA (tRNA) with methionine (Wallace *et al*, 2015). Therefore, protein aggregation can potentially occur without misfolding and exposure of hydrophobic regions.

### **3.2 Phase separation**

Another level of complexity to protein aggregation comes from the formation of phase separated membrane-less organelles. These structures involve accumulation of molecules to compartment that is not bound by membrane and actively exchanges material with its surrounding (Bolognesi *et al*, 2016). These dense droplets are formed from proteins that can be accompanied by nucleic acids (Nott *et al*, 2015).

In stress, proteins accumulate in phase separated organelles (Nott *et al*, 2015, Riback *et al*, 2017), such as stress granules formed in cytoplasm (Buchan & Parker, 2009) or stress bodies in nucleus (Biamonti & Vourc'h, 2010). Formation of phase separated organelles is a protective response that increases fitness upon stress (Riback *et al*, 2017). However, upon overexpression, accumulating proteins to phase separated organelles has also been shown to increase toxicity (Bolognesi *et al*, 2016).

Phase separated organelles are present in the cell also under normal conditions. These include, for example, nucleolus (Weber & Brangwynne, 2015) and P granules (Brangwynne *et al*, 2009). These findings highlight the many roles of phase separated organelles beyond protein deposit sites in stress. The driving force in the protein accumulation is probably not stemming from misfolding and exposed hydrophobic regions but rather from multivalent interactions (Boeynaems *et al*, 2018).

Proteins in phase separated organelles can in some conditions change their conformation and change the dynamics of the droplet completely (Alberti *et al*, 2017). For example, RNA-binding protein FUS form highly dynamic droplets under normal conditions which can develop to amyloid fibers when containing a disease mutation, changing the dynamic nature of the droplet to a rigid and toxic aggregate (Patel *et al*, 2015). This dynamic feature (in addition to rapid exchange of material with surroundings) makes phase separated organelles somewhat different from the classical view including amyloid fibers and amorphous aggregates.

### **3.3 Spatial aggregation**

During proteotoxic stress, all proteins around the cell are vulnerable for the harmful effects. Interestingly, cells have developed mechanisms to organize protein aggregates to specific cellular locations (Hill *et al*, 2017, Miller *et al*, 2015, Scior *et al*, 2016, Tyedmers *et al*, 2010). The benefits of spatial organization of aggregates include protection from toxic interactions between aggregates and other molecules, storage of the aggregates until more chaperones are available for them, selectively keeping the aggregates in one of the cells during cell division and it allows to organize protein quality control components to target specific aggregates (Miller *et al*, 2015). It should be noted that in addition to these organized aggregation sites, small aggregates not detectable by microscopic techniques might contribute to the aggregation load as well (Mogk *et al*, 2018).

In bacteria, misfolded proteins that overwhelmed quality control mechanisms are not able to handle are sequestered to one or two large inclusion bodies. The inclusion bodies are located at the poles of the bacteria (in case of two inclusion bodies) or at the center (in case of one inclusion body). The localization is thought to be driven by position of the nucleoid: one nucleoid in the center pushes the aggregates to the poles while duplicated nucleoids at each pole push the aggregates to the center. (Schramm *et al*, 2019)

Multiple protein aggregation sites have been described in yeast (Hill *et al*, 2017, Miller *et al*, 2015). Three main sites discovered in yeast are insoluble protein deposit (IPOD) close to vacuole, intranuclear quality control compartment (INQ) and cytosolic quality control-bodies (Kumar *et al*, 2017, Miller *et al*, 2015). While IPOD is proposed to sequester amyloidogenic and terminally aggregating proteins, INQ and cytosolic quality control-bodies are proposed to sequester protein aggregates formed in stress (Kumar *et al*, 2017, Miller *et al*, 2015). The sequestration to each aggregation site is to some extent conducted by specific collection of proteins (Hill *et al*, 2017, Miller *et al*, 2015). For example, small heat shock protein Hsp42 in yeast targets aggregates to peripheral deposit sites (Specht *et al*, 2011). It is thought that having multiple different aggregate deposit sites, yeast cells could arrange different quality control components to them. For example, ubiquitinated proteins destined for degradation would be sequestered to IPOD and cytosolic quality control-bodies (Hill *et al*, 2017, Miller *et al*, 2015). The protein quality control components could then be selectively arranged to each aggregation site (Hill *et al*, 2017).

In contrast to yeast, only one major aggregation site has been found in human cells. Aggresome is a cytoplasmic aggregation site covered with vimentin cage (Johnston *et al*, 1998). Proteins are sequestered to aggresomes when the ubiquitin mediated proteasomal degradation machinery is overwhelmed which leads to accumulation of misfolded proteins (Corboy *et al*, 2005). In addition to aggresome, other minor sites where proteins can accumulate in stress have been described. For example, misfolded



proteins are accumulated to nucleolus upon heat shock (Frottin *et al*, 2019). Although the nucleolus as a deposit site for misfolded proteins was found in human cells, it remains an open question whether it would have the same function in other eukaryotic organisms as well. Other protein deposit sites in human cells include nuclear stress bodies (Biamonti & Vourc'h, 2010) and cytosolic stress granules (Buchan & Parker, 2009), although they might have more functional role rather than a site for sequestering misfolded proteins.

## **4. Processing of protein aggregates**

Protein aggregates are not passively tolerated by cells. Wide range of cellular activities have developed to process the aggregates. Here, activities related to aggregate clearance are discussed. These include protein degradation, disaggregation and secretion.

### **4.1 Proteasomal degradation**

The ubiquitin-proteasome system is a protein degradation system that enables selective degradation of intracellular proteins. Proteins destined for proteasomal degradation are tagged with a chain of ubiquitin proteins. (Glickman & Ciechanover, 2002)

Ubiquitin is a 76 residues long evolutionary conserved protein. The process of tagging a target protein with ubiquitin is conducted by a set of enzymes called E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme) and E3 (ubiquitin-protein ligase). First, the ubiquitin is activated by E1 in an adenosine trisphosphate (ATP)-dependent manner to form a ubiquitin:E1 complex. The ubiquitin is then transferred to E2. Finally, the ubiquitin is transferred to  $\epsilon$ -NH<sub>2</sub> group (or N-terminal NH<sub>2</sub> group) of the target protein by E3. The E3 ligase can attach the ubiquitin straight from E2 to the target protein or it can involve an ubiquitin:E3 complex intermediate depending on the E3

ligase. The ubiquitin chain is recognized by a 26S proteasome which then degrades the target protein. (Glickman & Ciechanover, 2002)

The 26S proteasome is a barrel-like structure that is composed of 20S core particle and 19S regulatory particle. The 20S particle forms a hollow barrel-like structure. The proteolytic activity is located at the interior of the cavity. The 19S regulatory particle forms a cap-like structure to both end of the barrel-like 20S particle. While the 20S core particle degrades proteins, the 19S regulatory particle recognize the ubiquitin chain in a target protein and unfolds the protein to be inserted in the cavity of 20S particle for degradation. (Glickman & Ciechanover, 2002, Wolf & Hilt, 2004)

Ubiquitinated proteins are found in aggregated proteins. They can be found also on specific aggregation sites. For example, ubiquitinated proteins tend to be sequestered to INQ and cytosolic quality control-bodies in yeast rather than to IPOD (Hill *et al*, 2017, Miller *et al*, 2015). However, the presence of ubiquitinated proteins in aggregates might not mean they are destined for proteasomal degradation (Tyedmers *et al*, 2010); the ubiquitinated proteins in the aggregates could stem from overwhelmed proteasomal degradation system that is unable to process all its substrates (Glickman & Ciechanover, 2002, Tyedmers *et al*, 2010). Therefore, the sequestration of ubiquitinated proteins to aggregates could reflect a temporary storage place for irreversibly damaged proteins to be degraded later by proteasomes or clearance by autophagy (discussed later).

Proteasomes are able to degrade substrates that are unfolded (Wolf & Hilt, 2004). Therefore, proteasomal degradation of aggregated proteins would require an additional step of disaggregating the proteins first out from the aggregates. This can be achieved by coupling the proteasomal degradation system to chaperone-mediated disaggregation (discussed later). In nucleus - where the only protein degradation system is the proteasome - disaggregation and proteasomal degradation is coupled (Hjerpe *et al*, 2016). This coupling suggests, that proteasomal degradation has an important role in the nucleus that lack other protein degradation systems.

## 4.2 Autophagy

Autophagy is a catabolic process where cellular material is degraded in lysosomes. The substrates for autophagy range from protein aggregates to cellular compartments, such as mitochondria. Autophagy involves the formation of a membrane bound vesicle, termed autophagosome, that engulf material destined for degradation. It should be noted, that other processes under the term autophagy also exists that do not include formation of the autophagosome. The substrates engulfed by the autophagosome can be selected through adapter proteins or the process can be unselective. Eventually, autophagosome fuses with lysosome and the content of the autophagosome is degraded by lysosomal enzymes. (Galluzzi *et al*, 2017)

Protein aggregates can be degraded by autophagy, which is termed aggrephagy (Galluzzi *et al*, 2017). It involves adapter proteins such as HDAC6 (also linked to aggresome formation) and p62 that makes the process selective (Dikic, 2017, Johansen & Lamark, 2014). Ubiquitinated proteins are sequestered to aggresomes which are then cleared by autophagy (Dikic, 2017).

While aggrephagy is related to degradation of large aggregates, chaperone-mediated autophagy is used for degradation of soluble proteins. As with proteasomal degradation (discussed earlier), these processes require soluble proteins, i.e. disaggregation is required if the target proteins are from aggregates. Contrary to what their name suggests, chaperone-mediated autophagy does not include the formation of autophagosome. Instead, proteins are unfolded and transported directly to lysosomes by a channel formed in the lysosomal membrane. In addition to chaperone-mediated autophagy, chaperone-assisted selective autophagy is another similar mechanism for protein degradation in lysosomes. The two chaperone-linked mechanisms differ in the requirement of autophagosomal components. (Galluzzi *et al*, 2017)

### 4.3 Disaggregation

Disaggregation is the active process of re-solubilizing aggregated proteins. It is mediated by specialized chaperone systems (discussed later). After disaggregation, solubilized proteins can be refolded and continue delivering their function or irreversibly damaged proteins could be destined for degradation (Mogk *et al*, 2018). A proteome-wide study conducted with yeast indicate that proteins aggregated in heat shock are disaggregated during recovery without signs of degradation (Wallace *et al*, 2015). Therefore, disaggregation has an important role in cells recovering from proteotoxic stress.

In bacteria and non-metazoan (non-animal) eukaryotes, disaggregation is mediated by Hsp100 and Hsp70 chaperone systems. Together they form a disaggregase system that is able to re-solubilize aggregated proteins. (Doyle *et al*, 2013, Mogk *et al*, 2018, Parsell *et al*, 1994, Saibil, 2013, Sanchez & Lindquist, 1990).

The main component in the disaggregase is Hsp100 that forms an oligomeric ring-structure. The disaggregation is conducted by threading or pulling a substrate through a central tunnel in the ring structure (Saibil, 2013). The central tunnel contains loops that bind the substrate with aromatic residues and, upon ATP hydrolysis, carry the substrate a short distance through the tunnel (Mogk *et al*, 2018). This pulling is repeated by loops at different locations in the tunnel resulting in a substrate movement through the tunnel (Avellaneda *et al*, 2020).

The Hsp70 system assist Hsp100 in disaggregation by recruiting it to protein aggregates (Mogk *et al*, 2015). In addition, binding of Hsp70 activates the Hsp100 and enhances disaggregation (Mogk *et al*, 2015). After disaggregation, if the protein is re-folded, the Hsp70 system can also participate in this process. On the other hand,

metazoans (that do not have a Hsp100 outside mitochondria) are capable of disaggregation by Hsp70 system only (Nillegoda *et al*, 2018). The loss of Hsp100 (which is more powerful disaggregase than Hsp70) in metazoan could come from decreased aggregation load caused by the ability to move and, therefore, avoid proteotoxic environments (Nillegoda *et al*, 2018).

Hsp70 system is composed of Hsp70 that is assisted by Hsp40s (also known as J domain containing proteins or J proteins) and nucleotide exchange factors (Kampinga & Craig, 2010, Mayer, 2010). The main component is Hsp70 chaperone that has N-terminal ATPase domain and C-terminal substrate binding domain that contains a substrate binding pocket (Mayer & Gierasch, 2019). The substrate binding domain has also a lid structure that can cap the substrate binding pocket (Mayer & Gierasch, 2019). Upon ATP hydrolysis, the N-terminal domain induces a conformational change in the substrate binding domain that increases its substrate affinity (Mayer, 2010). Structurally speaking, the increase in affinity associates with closing of the lid on top of the substrate binding pocket (Mayer, 2010).

Hsp40s are a large group of proteins which all have an approximately 70 residues long J domain that has affinity for and stimulates ATPase activity of Hsp70 (Kampinga & Craig, 2010). Thus, they have important role in regulating the activity of Hsp70. In addition to Hsp70, some Hsp40s have affinity towards substrates (for example misfolded proteins) or certain cellular locations (for example membranes) (Kampinga & Craig, 2010). Therefore, many Hsp40s function also as adapter proteins bringing Hsp70 to substrates (or substrates to Hsp70) or to the needed location. The role of few Hsp40s in metazoan disaggregation has been established and so called canonical Hsp40s (class A and B in the protein family) have been recognized as guiding Hsp70s to protein aggregates (Nillegoda *et al*, 2017). Furthermore, Hsp40s have been shown to function as heterodimers in the context of disaggregation (Nillegoda *et al*, 2015). Different combinations of class A and B Hsp40s tune the Hsp70 system to disaggregate different types of aggregates (Nillegoda *et al*, 2015, Nillegoda *et al*, 2017).

Nucleotide exchange factors facilitate adenosine diphosphate (ADP) to ATP exchange for Hsp70 (Hartl *et al*, 2011). Hsp110 type nucleotide exchange factor - which interestingly is a homolog of Hsp70 (Mayer & Gierasch, 2019) - is essential for disaggregation in metazoa (Rampelt *et al*, 2012). In addition to Hsp40s, they regulate the activity of Hsp70; the ATPase activity and substrate binding is stimulated by Hsp40 while substrate release is stimulated by nucleotide exchange factors (Bracher & Verghese, 2015). Nucleotide exchange factors can also link Hsp70 systems to protein degradation and can have a role in determining the fate of the disaggregated substrate protein (i.e. whether it is refolded or degraded) (Bracher & Verghese, 2015).

#### **4.4 Aggregate secretion**

A straight-forward way to cope with protein aggregates is to secrete them out from to cell. The secretion can be done via different secretion pathways. Some of them result in aggregate secretion in lipid membrane containing vesicles while other secretion mechanisms do not package aggregates. While the aggregate secretion potentially reduces the proteotoxic burden of any particular cells, secreted aggregates are taken in by surrounding cells. Thus, the secretion (in multicellular organisms) can have dramatic effects at tissue or organism level. (Vaquer-Alicea & Diamond, 2019)

### **5 Protein aggregation and disease**

Many diseases are associated with the presence of protein aggregates. The most known examples include neurodegenerative diseases, such as Alzheimer's and Parkinson's disease. Typically, the pathological aggregates involve one kind of protein that is quite specific for each disease. For example, deposits of  $\alpha$ -synuclein are found in

the brains of Parkinson's disease patients. Other well characterized proteins found to aggregate in disease are TDP-43 in amyotrophic lateral sclerosis and huntingtin in Huntington's disease. (Chiti & Dobson, 2017)

In neurodegenerative diseases the causality between protein aggregation and disease development is not known. It has been proposed that three links exist. First, the formation of protein aggregates is toxic and contributes to the disease development. Second, protein aggregation is one outcome of the disease and follows disease development. Third, protein aggregation is a protective mechanism to sequester toxic proteins. The toxicity related to protein aggregation might not be related to the larger aggregates detectable with microscopy techniques. Rather, the toxic aggregates are the small oligomers of misfolded proteins. (Espay *et al*, 2019)

Interestingly, all the discussed diseases have the protein deposits at cells from brain tissues. The low protein quality control capacity in brain tissues is probably an important factor (Kundra *et al*, 2020). Pathological protein aggregation is not, however, limited to brain. Diseases related to protein aggregation can originate, for example, in pancreas (diabetes) or eye (cataracts) (Valastyan & Lindquist, 2014).

## **6. Mass spectrometry-based proteomics in protein aggregation studies**

Mass spectrometry (MS) is a widely used technique in biology. Its applications range from detecting and quantifying all major biomolecules: lipids, carbohydrates, nucleic acids, metabolites and proteins. The major advantage of MS is its capability to analyze multiple targets simultaneously. For example, a typical MS-based proteomics experiment includes analysis of thousands of proteins. Recent developments in MS

have made the technique more available for non-specialist users with reduced costs. At the same time, the quality and capabilities of MS instruments have increased. (Cox & Mann, 2007, Griffiths & Wang, 2009)

## **6.1 Principles of mass spectrometry**

The core principle in any MS technique is the analysis of mass to charge ratio of charged particles. Two essential requirements are a particle with a charge and a vacuum where the movement of that charged particles can be controlled. The mass to charge ratio can be determined by various techniques. The most common ones include time of flight and orbitrap. In time of flight -type MS, the mass to charge ratio is determined from a time it takes for a charged particle to travel a certain distance in a specific magnetic field. The travel time is proportional to charge and mass of the particle. In orbitrap type MS, the analyzed particles are set to an oscillating motion and the mass to charge ratio can be analyzed from the frequency of electromagnetic radiation send by the particles; during oscillations, the particles accelerate and decelerate creating electromagnetic waves with frequencies that are specific for their mass to charge ratio. (Aebersold & Goodlett, 2001, Domon & Aebersold, 2006, Eliuk & Makarov, 2015, Scigelova *et al*, 2011)

## **6.2 Mass spectrometry-based proteomics**

Protein analysis by MS follows the same principles as described in the previous section. Generally, two different kinds of proteomics approaches exist: top-down and bottom-up. In top-down technique, typically a full length protein is injected to MS and the mass to charge ratios of the whole protein is analyzed. The application for top-down proteomics include protein mass determination, for example. In bottom-up proteomics, proteins are first digested to small peptides with sequence specific protease (often with trypsin that



cleaves proteins after lysine and arginine residue) which are then injected to MS for analysis. Peptides are analyzed first as intact peptides (so called MS1 scan) and then again after fractionation (so called MS2 scan) that takes place in the MS instrument. The MS1 scan analyzes the mass to charge ratio of the peptide. After MS1 scan, the peptides are fragmented. During fragmentation, the weakest covalent bonds in peptide - typically the peptide bond - breaks. The fragmentation results in a pool of fragmented peptides. In MS2 scan, the mass to charge ratios of these fragments of the parent peptide are analyzed. In case peptide level labelling was used (discussed later), the labels are also detected and quantified in MS2 scan. (Gillet *et al*, 2016)

The outcome of MS2 analysis is spectra that includes intensities and mass to charge ratios of ions created during the fragmentation. The spectra are compared to a spectra created *in silico* for all chosen proteins. Typically, the proteins are selected from the proteome of the analyzed organisms. The peptide identity can be analyzed by comparing the experimentally and *in silico* obtained spectra. The comparison also allows to determine the proteins from which the peptide originated. Therefore, each analyzed peptide can be connected to a certain protein. (Gillet *et al*, 2016)

### **6.3. Quantitative mass spectrometry**

Traditionally, mass spectrometry-based proteomics have been used to identify proteins. Recent developments include quantification methods that, when combined with the protein identification analysis, allow quantitative measurement of large number of proteins simultaneously. Quantitative methods can be classified to label-free and labelling methods. Here, only labelling methods are discussed. Labelling methods can be further divided to protein level and peptide level labeling. (Bantscheff *et al*, 2012)

### 6.3.1 Protein level labelling by SILAC

One of the most utilized protein level labelling method is stable isotope labelling of amino acids in cell culture (SILAC). In SILAC, cells (or even multicellular organisms, such as mice) are fed with amino acid supply that contains stable isotopes of nitrogen or carbon atoms in lysine or arginine residues. Cells adopt these amino acids to proteins during protein synthesis. By having different isotope composition for different samples, e.g. cell culture flasks, SILAC allows to distinguish these samples during MS analysis. (Bantscheff *et al*, 2012, Ong *et al*, 2002)

Since in SILAC the labelling is done for proteins, the samples can be pooled on a very early stage (Bantscheff *et al*, 2012). This decreases the error rates (stemming from pipetting, for example) as compared to peptide level labelling where the pooling can be done much later in the sample processing.

Different experimental usages of SILAC has been described. The simplest approach is to quantitatively compare the proteins from two or more samples (Ong *et al*, 2002). For example, drug treatment and control. Further developments of SILAC include pulsed labelling where the growth medium (including the modified and isotopically labelled amino acids) is changed instead of comparing two fully SILAC labelled samples (Fierro-Monti *et al*, 2013). This design allows to follow the incorporation of the amino acids from the newly added medium. At the same time, removing the old medium means that proteins do not have a bulk supply of that particular SILAC label. The pulsed design is often used to quantify protein turnover i.e. synthesis and degradation (Doherty *et al*, 2009, Fierro-Monti *et al*, 2013, Mathieson *et al*, 2018, Zecha *et al*, 2018).

### 6.3.2 Peptide level labelling by tandem mass tags

While protein level labelling is conducted with intact proteins (and usually in cell culture with intact cells), peptide level labelling takes place after cell lysis and protein digestion with sequence specific protease (Bantscheff *et al*, 2012). Multiple peptide labelling methods exist (Chahrour *et al*, 2015) but only tandem mass tags (TMT) is discussed here.

TMT labels are set of molecules with the same overall mass that are attached to primary amines in peptides (lysine residues and N-terminus). Therefore, a peptide labelled with a TMT label has the same mass as the same peptide from another sample labelled with a different TMT label; after pooling the samples, they are not distinguished from each other in the MS1 scan. However, during the fragmentation prior MS2 scan, the TMT label breaks into two parts. One stays intact with the peptide while the another one (so called reporter ion) is cleaved off. The carbon and nitrogen isotopes are arranged differently in each TMT label so that - although having the same overall mass - after the cleavage, the reporter ions of each TMT have different mass. Similarly, the mass difference is compensated by the mass of the label that stayed intact to the peptide. The reporter ions are also analyzed during MS2 scan. The intensity of the reporter ions (relative to the other reporter ions) allows comparative quantification of the peptide between different samples. (Thompson *et al*, 2003)

A separation of 1 Da between the masses of reporter ions (of different TMT labels) is achieved by different number of neutrons in each reporter ion (Thompson *et al*, 2003). Furthermore, an additional separation within reporter ions is achieved by adding a neutron on either nitrogen or carbon atom (Werner *et al*, 2014). The difference in the energy of strong interactions of the two different nucleoids is enough to cause a mass shift (following Einstein's formula  $E = mc^2$ ) that can be measured with modern mass

spectrometers (approximately 6 mDa). At the moment, the highest multiplexing capacity of the type of TMT labels describe here is 16 samples (Li *et al*, 2020).

### **6.3.3 Hyperplexing - combination of labelling techniques**

Some labelling techniques, such as SILAC and TMT, label quite different parts of the protein or peptide. Therefore, the combination of the two has been used to increase the multiplexing capacity, termed hyperplexing (Dephoure & Gygi, 2012). The most used application take advantage of SILAC to label and differentiate mature from newly synthesized proteins and TMT for multiplexing different time points (Aggarwal *et al*, 2019). This has allowed to analyze the dynamics of protein turnover with high temporal resolution (Savitski *et al*, 2018, Zecha *et al*, 2018).

## **6.4 Mass spectrometry in protein aggregation studies**

Mass spectrometry-based proteomics has been used to study protein aggregation in many model organisms under various conditions. For example, with yeast, the aggregation has been studied in the context of heat shock (O'Connell *et al*, 2014, Wallace *et al*, 2015), exposure to arsenite (Ibstedt *et al*, 2014) and under various chemical stresses (Weids *et al*, 2016). Importantly, these studies are also conducted under normal physiological conditions to find that some proteins form insoluble deposits in the absence of proteotoxic stress in yeast (Ibstedt *et al*, 2014, Weids *et al*, 2016) as well as in human cells (Sridharan *et al*, 2019), for example. The nematode *C.elegans* has been used in aging related proteome-wide aggregation studies (David *et al*, 2010, Vecchi *et al*, 2019, Walther *et al*, 2015). Aggregation studies done with non-human mammals include *in vivo* aggregate analysis from mice that express aggregation-prone huntingtin protein (Hosp *et al*, 2017). In addition, aggregation was monitored after various stress conditions using mice cells (Sui *et al*, 2020).

At the time of writing, only one study done with yeast has been published where protein disaggregation was analyzed with proteomics (using SILAC quantification) (Wallace *et al*, 2015).

## **AIMS AND OBJECTIVES**

### **1. Aims**

The aim of the project was to study how human proteome responses to proteotoxic stress. The proteotoxic stress would be transient exposure of cells to non-lethal elevated temperature. The cellular response under investigation would be narrowed to protein solubility, protein synthesis and thermal stability.

The response would be studied from two perspectives. First, what would be the immediate response to proteotoxic stress. Second, what would happen to the proteome during recovery from the stress. Especially, how would the cells respond to the immediate stress-induced changes such as protein aggregation.

The main aim of the project was to study heat-induced protein aggregation of endogenous proteins in human cells and their potential disaggregation during recovery. At the time of writing, no previous reports related to these topics on proteome-wide scale exist. Therefore, after completion, the main aim would fill an important gap in the literature. The secondary aims of the project were heat shock-induced effects on protein synthesis and on thermal stability of proteins that would not aggregate upon heat shock. The experiment designed to study the main aim allows also to study protein synthesis. Studying thermal stability upon heat shock involves the application of novel mass spectrometry technologies to the context of proteotoxic stress and would results in a unique dataset for creating new hypothesis.

## **2. Objectives**

### **2.1 Development of mass spectrometry-based platform to study protein solubility after heat shock and during recovery**

To achieve this objective, a method was developed that would have the following principles: (1) quantitative proteome-wide analysis, (2) protein solubility measurement and (3) ability to sample multiple time points. The first principle was achieved by using bottom-up mass spectrometry-based proteomics. The second principle was covered by utilizing weak detergents that allows to separate soluble proteins from insoluble ones. The third principle was achieved by utilizing multiplexing methods.

### **2.2 Characterization and quantification of proteins prone for aggregation in heat shock**

The aggregating proteins would be identified based on the solubility change upon heat shock. After robust statistical testing, identity of aggregating proteins could be determined.

To characterize aggregation prone proteins, statistical analysis of different protein features was applied. Relevant features under investigation were chosen based on the existing literature of proteome-wide aggregation studies conducted with non-human model organisms, such as yeast, nematode and mice. This would also allow species-wide comparison of proteins aggregation.

### **2.3 Characterization and quantification of protein disaggregation during recovery from heat shock**

The disaggregation of aggregating proteins would be analyzed from the changes in protein solubility during recovery from heat shock. Multiple time points sampled in the recovery phase allows to have a high temporal resolution and makes dynamic measurement of the disaggregation possible.

### **2.4 Analysis of heat shock-induced changes in protein synthesis**

The multiplexing used in the developed method included labelling of newly synthesized proteins. Therefore, the protein synthesis can be quantified from the same data used for estimating protein solubility. The labelling of newly synthesized proteins would result in accumulation of the label and allows to approximate the synthesized protein amounts.

### **2.5 Analysis of heat shock-induced changes in thermal stability of non-aggregating proteins**

Thermal stability of proteins that do not aggregated on heat shock would be estimated upon heat shock. This is possible by utilizing two-dimensional thermal proteome profiling.



## MATERIALS AND METHODS

### List of reagents and resources

Reagent/resource	Reference or source	Identifier or catalog number
<b>Experimental models</b>		
K-562 human chronic myelogenous leukemia cell line	ATCC	CCL-243
<b>Chemicals, enzymes and other reagents</b>		
SILAC RPMI 1640 medium	Thermo Fisher Scientific	88365
L-glutamine	Sigma-Aldrich	G7513
L-lysine (light)	Thermo Fisher Scientific	89987
L-arginine (light)	Thermo Fisher Scientific	89989
Dialyzed fetal bovine serum (FBS)	Thermo Fisher Scientific	26400044
L-lysine (heavy) $^{13}\text{C}_6^{15}\text{N}_2$	Thermo Fisher Scientific	88209
L-arginine (heavy) $^{13}\text{C}_6^{15}\text{N}_4$	Thermo Fisher Scientific	89990

PBS	In-house service	
HEPES	In-house service	
cOmplete EDTA-free Protease inhibitor cocktail	Roche	11873580001
PhosSTOP (phosphatase inhibitor cocktail)	Roche	04906845001
NP-40 (IGEPAL CA-630)	Sigma-Aldrich	18896
MgCl <sub>2</sub>	In-house service	
Benzonase Nuclease	Millipore	71206-3
SDS	Bio-Rad	161-0418
Ethanol	Sigma-Aldrich	34852
Formic acid	Biosolve	069141
Water, LC-MS grade	Fisher Chemical	W6-212
Acetonitrile, LC-MS grade	Fisher Chemical	A955212
Hydroxylamine	Sigma-Aldrich	438227
TMT10	Thermo Fisher Scientific	90111
TMT11	Thermo Fisher Scientific	A37724
Chloroacetamide	Sigma-Aldrich	C0267
Tris(2- carboxyethyl)phosphine hydrochloride	Sigma-Aldrich	C4706

Trypsin	Promega	V5111
LysC	FUJIFILM Wako	125-05061
DMSO	Sigma-Aldrich	276855
Ammonia	Sigma-Aldrich	221228
	Sigma-Aldrich	78314
Buffer A (0.1% formic acid in water)		
Buffer B (0.1% formic acid in acetonitrile)		
Running buffer (10x Tris/Glycine/SDS)	Bio-Rad	1610732
Coomassie stain (Coomassie Brilliant Blue R-250 Dye)	Thermo Fisher Scientific	20278
SDS-PAGE gel (10% Criterion Tris-HCl Protein Gel)	Bio-Rad	3450009
DTT (Dithiothreitol)	Sigma-Aldrich	R0862
<b>Software</b>		
isobarQuant	(Franken <i>et al</i> , 2015) <a href="https://github.com/protcode/isob">https://github.com/protcode/isob</a>	
Mascot	Matrix Science <a href="http://www.matrixscience.com/">http://www.matrixscience.com/</a>	

R	R Core Team  <a href="https://www.R-project.org">https://www.R-project.org</a>	
---	--	--

<b>Databases</b>		
UniProt	(Consortium, 2019)  <a href="https://www.uniprot.org/">https://www.uniprot.org/</a>	
Human Protein Atlas	(Thul <i>et al</i> , 2017)  <a href="http://www.proteinatlas.org">http://www.proteinatlas.org</a>	
Protein complexes	(Ori <i>et al</i> , 2016)  <a href="http://www.bork.embl.de/Docu/variable_complexes/">http://www.bork.embl.de/Docu/variable_complexes/</a>	
Database of Disordered Protein Predictions	(Oates <i>et al</i> , 2013)  <a href="http://d2p2.pro">http://d2p2.pro</a>	
Gene Ontology Annotation Database	(Huntley <i>et al</i> , 2015)	
<b>Other</b> (commercial kits and consumables)		
96-well PCR plates	Eppendorf	0030133366
Stericup 0.22 µm filter	Merck Millipore	S2GPU01RE
Aluminum cover foil	Thermo Fisher Scientific	AB0626
Vent filter membrane	VWR	60941-086

CellTiter-Glo Luminescent Viability Assay	Promega	G7571
Optiplate-96 Luminescence plate	PerkinElmer	
0.2 ml strip tubes	ratiolab	8610040
Microcentrifuge	Carl Roth	
Table centrifuge (Multifuge X3R)	Thermo Fisher Scientific	
Liquid nitrogen		
Shaker (Thermomixer Comfort)	Eppendorf	
Filter plates (0.45 µm)	Merck Millipore	MSHVN4550
Filter plates (0.22 µm)	Merck Millipoer	MSGVN2250
BCA Protein Assay Kit	Thermo Fisher Scientific	23225
Carboxylate modified magnetic particles, hydrophilic	Sigma-Aldrich	45152105050250
Carboxylate modified magnetic particles, hydrophobic	Sigma-Aldrich	65152105050250
OASIS HLB µElution plate	Waters	186001828BA
96-well microplates	Thermo Fisher Scientific	249944

96-well microplates (conical bottom)	Greiner Bio One	651201
Trapping cartridge. Acclaim PepMap 100 C18 LC column; 5 µm particles with 100 Å pores; 5 mm column with 300 µm inner diameter	Thermo Fisher Scientific	
Analytical column. nanoEase HSS C18 T3, 75 µm x 25 cm, 1.8 µm, 100 Å	Waters	
<b>Devices</b>		
UltiMate 3000 RSLC nano LC system	Thermo Fisher Scientific	
Q Exactive Plus Orbitrap mass spectrometer	Thermo Fisher Scientific	
Orbitrap Fusion Lumos mass spectrometer	Thermo Fisher Scientific	
SureCycler 8800 Thermal Cycler	Agilent	
Infinite M1000 PRO plate reader	TECAN	
TC20 cell counter	Bio-Rad	

1290 Infinity (for HPLC fractionation)	Agilent	
Vacuum concentrator Univapo 150 ECH	UniEquip	

## 1. Cell culture

K-562 cells were maintained in light medium at T25 or T75 flasks in a cell culture incubator [37°C, 5% carbon dioxide (CO<sub>2</sub>)]. The light medium contained SILAC RPMI 1640 medium supplemented with 2 mM L-glutamine, 0.96 mM L-lysine (light), 0.48 mM L-arginine (light) and 10% dialyzed fetal bovine serum (FBS), all filtered through a 0.22 µm Stericup filter. Cells were diluted with fresh and pre-warmed (37°C) medium every two to three days to maintain the cell density between 5x10<sup>5</sup> - 1x10<sup>6</sup> cells/ml.

## 2. Preparing cells for heat treatments

For dynamic SILAC experiment, the light medium was switched to heavy medium 90 minutes before heat treatments. The heavy medium contained same components and concentrations as the light medium, except heavy L-lysine and L-arginine were used. These heavy versions of the amino acids contained heavier stable carbon and nitrogen isotopes increasing their mass by eight (L-lysine) or ten (L-arginine) daltons.

For the medium switch, cells were pelleted with centrifugation [190 x g, 3 min, room temperature (RT)], supernatant was removed, and the remaining cell pellet was gently re-suspended to pre-warmed (37°C) heavy medium. Cell were pelleted again,

supernatant was removed and cells were re-suspended to pre-warmed heavy medium with to a final cell density of  $5 \times 10^5$  cells/ml. Cells were kept in cell culture incubator for 90 minutes to allow for the consumption of all residual light labelled amino acids.

After the medium switch, all mature pre-existing proteins in the cells have light version of L-lysine or L-arginine, while all newly synthesized proteins adapt the heavier version of them. This allows to later separate peptides originating from these two protein fractions from each other in MS measurements.

To prepare cells for two dimensional thermal proteome profiling (2D-TPP) experiment, where dynamic SILAC was not applied, cells maintained in light medium were pelleted with centrifugation (190 x g, 3 min, RT), supernatant was removed and cells were re-suspended to fresh pre-warmed (37°C) light medium to a final cell density of  $1.5 \times 10^6$  cells/ml.

The different final cell density between dynamic SILAC and 2D-TPP experiments, although the heat treatments are conducted the same way, is explained by the high number of cells required for 2D-TPP experiment. At the same time, the space used in heat treatment (96 wells/treatment) was limited. Therefore, the final cell density in the heat treatments in 2D-TPP experiment was higher than in the dynamic SILAC experiment.

Experiments were conducted as biological triplicates and each replicate was prepared on different day. However, in 2D-TPP experiment, only two replicates of mock shocked samples were prepared due to limitations in sample arrangement in TMT10 labeling workflow (see Figure 30 for the sample arrangement).



### **3. Heat treatment**

The heat shock was initially conducted on water bath in the pilot experiments. However, in the main experiment, the heat treatments were conducted with heat blocks.

#### **3.1 Water bath**

Water bath was pre-heated to 45°C. Heavy medium was pre-warmed in the water bath. Before heat shock, cells were transferred to a 50 ml conical tube, pelleted (190 x g, 3 min, RT), supernatant was removed and cells were re-suspended to 10 ml of pre-warmed heavy medium. Cells in the conical tube were immersed in the water bath for ten minutes. During this heat treatment, cells were constantly kept in motion manually to ensure smooth distribution of temperature in the sample. After heat shock, cells were immersed in ice bath for 30 seconds to cool. Cells were let to recover in cell culture incubator (37°C, 5% CO<sub>2</sub>) for 0, 5, 10, 15, 20, 30, 40, 60 or 120 minutes before collecting samples. Sample collection was conducted as described later.

#### **3.2 Heat block**

Cells were aliquot to two 96-well PCR plates. Each well contained 200 µl of cell suspension corresponding to 100 000 cells/well in dynamic SILAC experiment and 300 000 cells/well in 2D-TPP experiment. The plates were sealed with aluminum cover foils.

The heat treatments were conducted in pre-heated heating blocks. One plate was put in a heating block at 37°C (mock shock) and another plate in a heating block at 44°C (heat shock). The plates were kept in the heating block for ten minutes.

For dynamic SILAC experiment, the aluminum cover foils were replaced with vent filter membranes after the heat treatment. The plates were put back to cell culture incubator (37°C, 5% CO<sub>2</sub>) to allow cells to recover from the heat treatment.

For 2D-TPP, the aluminum cover foils were removed, cells were pooled to and prepared for 2D-TPP denaturation assay (described later).

#### **4. Sample collection**

Samples in experiment with heat shock and recovery were collected after heat shock and after one, two, three and five hours of recovery. Samples were transferred to a 0.2 ml strip tubes. Cells were pelleted with microcentrifuge (100 x g, 1 min, RT), 90% of supernatant (180 µl) was removed and cells were re-suspended to 180 µl of ice cold phosphate-buffered saline (PBS). The pelleting and removal of supernatant was repeated. Cell pellets with the residual 10% (20 µl) of PBS were snap-frozen in liquid nitrogen.

Samples in experiment with heat shock and recovery aiming to analyze the total protein amount were collected as described above. The exception was that no samples after two hours of recovery were collected.

#### **5. Viability assay**

The effects of different heat shock temperatures on cell viability was estimated with cell viability assay (CellTiter-Glo Luminescent Viability Assay). Cells cultured in light medium were pelleted (190 x g, 3 min, RT). The pellet was re-suspended to light medium to allow adjustment of the cell density to  $5 \times 10^4$  cells/ml. Cells were distributed to 96-well microplates. Each well contained 200 µl of cell suspension corresponding to  $10^4$  cells/well. Two plates were prepared for samples to be collected right after heat shock and five hours after recovery. The plates were sealed with aluminum cover foil.

For heat shock, each plate was put on a heating block (SureCycler 8800 Thermal Cycler) programmed to have a heat gradient of temperatures 37.0; 39.6; 41.5; 43.6; 45.7; 47.5; 49.6; 52.0; 54.3 and 54.9°C. Cells were treated for ten minutes in the heat block.

After heat shock, the aluminum cover foil was removed from the plate. The plate containing samples to be measured after five hours of recovery, was sealed with vent filter membrane and placed in a cell culture incubator (37°C, 5% CO<sub>2</sub>).

For the viability assay, samples were gently mixed with pipette and half (100 µl) of each sample was transferred to a luminescence plate (Optiplate-96). The remaining samples on the microplate were pelleted (190 g, 3 min, RT) and supernatant transferred to the luminescence plate for matching blank controls (to control for possible heat-induced changes in the medium). For each sample on the luminescence plate, equal volume of CellTiter-Glo working solution (prepared according to the manufacturer's instructions) was added. The plate was mixed on a shaker (750 rpm, RT) for two minutes and the reaction was let to settle for ten minutes without shaking. Luminescence was measured with a plate reader (Infinite M1000 PRO).

For each sample, the luminescence of the blank was subtracted and a ratio against heat shock at 37°C was calculated. Each sample was prepared as technical triplicates.

## **6. Two dimensional proteome profiling**

2D-TPP (Becher *et al*, 2016), an extension of thermal proteome profiling (Franken *et al*, 2015, Mateus *et al*, 2017, Reinhard *et al*, 2015, Savitski *et al*, 2014), was conducted to compare changes in thermal stability of soluble proteins upon heat shock. Cells that had undergone heat or mock shock were pelleted (180 x g, 3 min, RT), pellets re-suspended to PBS and pelleted again. Pellets were re-suspended to PBS to a final cell density of  $5.5 \times 10^6$  cells/ml. Cells were distributed on a 96-well microplate. Each well contained

100 µl of cell suspension corresponding to  $5.5 \times 10^5$  cells/well. Cells were pelleted on the plate (390 x g, 2 min, RT) and 80 µl of the supernatant was removed. The pellets were gently re-suspended to the remaining 20 µl of supernatant.

The plate was placed in a thermal cycler (SureCycler 8800) programmed to have a temperature gradient of temperatures 37.0; 37.8; 40.4; 44.0; 46.9; 49.8; 52.9; 55.5; 58.6; 62.0; 65.4 and 66.3°C. The cells were treated with the heat gradient for three minutes and then let to settle in RT for another three minutes before placing the plate on ice.

## **7. Cell lysis**

For experiments with heat shock and recovery, cells were thawed on ice. Cells were lysed by mixing them with 30 µl of ice cold lysis buffer prepared as 5/3x concentrate. The lysis buffer (when diluted to 1x) contained 50 mM HEPES, 0.8% NP-40, 1.5 mM  $MgCl_2$ , 1x protease inhibitor cocktail, 1x phosphatase inhibitor cocktail and 0.25 U/µl Benzonase. Samples were kept at 4°C on a shaker for one hour.

2D-TPP samples were lysed as above without the need to thaw the cells.

Samples estimating the total protein amount were lysed as above except NP-40 was replaced by 1% sodium dodecyl sulfate (SDS). After mixing cells with the lysis buffer, samples were kept at RT on a shaker for 30 minutes.

## **8. Removal of insoluble fraction**

Wells in 0.45 µm filter plate were pre-wet with 1x lysis buffer without Benzonase. Excess buffer was removed by centrifugation (1200 rpm, 2 min, 4°C). Lysates were

spun down and transferred to the filter plate. Soluble fractions were collected on a 96-well microplate by centrifugation (500 g, 5 min, 4°C) while the insoluble fraction with cell debris was captured on the filter plate.

Soluble fractions were either dried, frozen or processed immediately.

The insoluble fraction on the filter plate was washed by addition of lysis buffer to the filter plate and centrifugation (500 x g, 5 min, +4 °C). The washing was repeated. Insoluble fraction was brought to solution by adding 2% SDS to the filter plate and incubating the plate at 56°C for 30 minutes. The insoluble fraction was collected by centrifugation (500 x g, 5 min, +4 °C).

Protein concentrations in the soluble fraction were analyzed by BCA protein assay kit according to manufacturer's instructions. For protein concentration assay, samples were prepared as 1/4 dilutions in water.

## **9. Protein extraction**

Samples were thawed if necessary and SDS was added to a final concentration of 1%.

Proteins were extracted using a modified version of single-pot solid-phase-enhanced sample preparation (Hughes *et al*, 2014, Moggridge *et al*, 2018). Samples containing 5-15 µg of proteins were transferred to a 96-well microplate. 1% SDS was added to level the sample volumes.

Samples were mixed with carboxylate modified magnetic particles (1:1 mixture of hydrophobic and hydrophilic) in 47.6% ethanol and 2.4% formic acid on a shaker at RT for 15 minutes to bind proteins to the particles.

Samples were transferred to a 0.22 µm filter plate and non-bound material was removed from the particles by centrifugation (1000 x g, 1 min, RT). Particles were washed with 200 µl of 70% ethanol followed by a centrifugation (1000 x g, 2 min, RT) to remove the ethanol. The ethanol wash was repeated three times.

## **10. Tryptic digestion**

Particle-bound proteins were mixed with digestion mixture containing 90 mM HEPES, 5 mM chloroacetamide, 1.25 mM tris(2-carboxyethyl)phosphine, 200 ng trypsin/sample and 200 ng LysC/sample. The digestion was conducted on a shaker at RT for overnight. Parallel to digestion, proteins were eluted out from the particles into the aqueous buffer.

After digestion, peptides were collected by centrifugation (1000 x g, 1 min, RT). To collect any residual peptides that remained bound to particles, 10 µl of 2% dimethyl sulfoxide was added and, after centrifugation (1000 x g, 1 min, RT), flow-through was collected and pooled with the previously eluted peptides. Samples were dried in a vacuum concentrator.

## **11. TMT labeling**

Dried tryptic peptides were dissolved in water. TMT-labels were dissolved in acetonitrile and mixed with the peptides with a final acetonitrile concentration of 28.6%. The labeling reaction was let to develop at RT on a shaker for one hour. The reaction was quenched by adding hydroxylamine (with final concentration of 1.1%). The quenching was done at RT on a shaker for 15 minutes.

Labelled samples were pooled and diluted with 0.05% formic acid to lower the acetonitrile concentration below 5% needed for peptide binding to desalting columns.

## **12. Peptide desalting**

Desalting plates (OASIS HLB  $\mu$ Elution plate) were primed by washing the columns with 0.05% formic acid / 80% acetonitrile. Samples were cleaned from any residual particles from protein extraction step by placing them on a magnetic rack. Supernatants were transferred to the desalting plate. Peptides were bind to the columns by flushing the samples through the columns with vacuum. Column-bound peptides were washed twice with 0.05% formic acid. Peptides were eluted with 0.05% formic acid / 80% acetonitrile to collection vials and dried in a vacuum concentrator.

## **13. Off-line fractionation**

Dried peptides were dissolved in 20 mM ammonia. Peptides were separated according to hydrophobicity with reversed-phase chromatography under high pH conditions. From the separation, 32 fractions were collected on a 96-well microplates (with conical bottom). Fractions from the beginning (first) and end of the gradient (two last fractions) were omitted. The remaining fractions were pooled to gain 12 fractions. The pooled fractions were chosen from distant locations of the gradient (at least ten fractions away from each other). The fractions were dried in a vacuum concentrator.

## **14. Quantitative mass spectrometry**

Fractionated peptides were dissolved in 0.1% formic acid. Peptides were subjected to liquid-chromatography (LC) (UltiMate 3000 RSLC nano LC system) attached to mass spectrometer. Peptides were loaded on a trapping cartridge for desalting (3 min with 0.05% trifluoroacetic acid with flow rate of 30  $\mu$ l/min). Peptides were eluted with buffer A and B with increasing concentrations of buffer B (0.3  $\mu$ l/min) to an analytical column. Concentration of buffer B increased with the following steps: from initial 2% to 4% in the

first four minutes, from 4% to 8% in the next two minutes, from 8% to 28% in the next 96 minutes and from 28% to 40% in the next ten minutes. Finally, a wash with 85% buffer B was conducted for four minutes before returning to initial conditions. The total analysis time was 120 minutes.

From the LC, peptides were injected to mass spectrometer. Analysis of samples from heat shock and recovery experiments was conducted with Q Exactive Plus Orbitrap while samples from 2D-TPP experiment were analyzed with Orbitrap Fusion Lumos.

Q Exactive Plus Orbitrap was operated in positive ion mode with 2.3 kV spray voltage and 275°C capillary temperature. Full scan MS spectra were acquired for a mass range of 375-1200 m/z in profile mode with a resolution of 70,000 (maximum fill time of 250 ms or automatic gain control with a maximum of  $3 \times 10^6$  ions). Data-dependent acquisition was applied on the MS1 scan by fragmenting top ten peptide peaks (cycle time of 3 seconds) with charge state between 2-4. For isolation, dynamic exclusion window of 30 seconds and mass window of 0.7 m/s was used. Selected peptides were then fragmented using normalized collision energy of 32. For MS2 spectra acquisition, a resolution of 35,000 and automatic gain control target of  $2 \times 10^5$  ions were used in profile mode. The first mass was set to 100 m/z.

Orbitrap Fusion Lumos was operated in positive ion mode with 2.4 kV spray voltage and 300°C capillary temperature. Full scan MS spectra were acquired for a mass range of 375-1500 m/z in profile mode with a resolution of 120,000 (maximum fill time of 64 ms or automatic gain control with a maximum of  $4 \times 10^5$  ions). Data-dependent acquisition was applied on the MS1 scan by fragmenting top ten peptide peaks (cycle time of 3 seconds) with charge state between 2-7. For isolation, dynamic exclusion window of 60 seconds and mass window of 0.7 m/s was used. Selected peptides were then fragmented using normalized collision energy of 38. For MS2 spectra acquisition, a resolution of 30,000 and automatic gain control target of  $1 \times 10^5$  ions were used in profile mode. The first mass was set to 100 m/z.



## 15. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

Disulphide bonds in proteins were reduced by adding 5 µl of 200 mM DTT to each sample. Samples were mixed and incubated at +95 °C for 5 minutes to denature proteins. Samples were loaded on a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel. Proteins were separated first with 80V for 20 minutes and then with 200V for 35 minutes in running buffer. The gel was stained for two hours in staining solution (60 mg/ml Coomassie stain in 20% ethanol solution), destained with 20% ethanol for several times and scanned.

## 16. Data analysis

Raw data acquired from the MS analysis was processed with *preMascot* workflow in isobarQuant (Franken *et al*, 2015) analysis package to pre-process the raw data and create a peak list for peptide and protein identification with Mascot.

Peptide and protein identification was conducted with Mascot search engine. Peaks from MS/2-scans (i.e. intensities of ions from fragmented peptides) were compared to *in silico* fragmented tryptic peptides originating from human proteome [UniProt (Consortium, 2019)] to identify measured peptides. The search was conducted for tryptic peptides (cleavage after C-terminal lysine or arginine) with error tolerance of 10 ppm for peptide mass and 0.02 Da for fragment ions in MS/MS. In addition, carbamidomethylation of cysteines and TMT labelling of lysines was required for all peptides (fixed modifications) and acetylation of N-terminus, methionine oxidation and TMT labelling of N-terminus were allowed (variable modification).

For experiments utilizing SILAC labelling, the search was conducted twice for each fractions (light and heavy). The search was done as described above for the light fraction. For heavy fraction, requiring simply a heavier lysine and arginine in the search space is not possible since Mascot allows only for one modification for each amino acid; each lysine would be required to contain a TMT label and heavier stable isotopes. This

issue was circumvented by creating a special lysine-attached TMT label that would contain the mass of a TMT label and the extra mass (8 Da) of heavy lysine.

After peptide and protein identification, the data was further processed with *postMascot* workflow of the isobarQuant (Franken *et al*, 2015) package to add all peptide level data to protein quantification (with quality control measures).

The protein level data was imported to R and cleaned by removing proteins identified as reverse database hits (in the Mascot search) or contaminants (such as bovine serum albumin used in instrument calibrations). To increase the quality of data, only proteins that were quantified with at least two peptides unique to that protein were kept.

Protein intensities in each TMT channel were log2-transformed. Possible batch effects were removed with *removeBatchEffects* function in R package *limma* (Ritchie *et al*, 2015). Intensities were normalized with variance stabilization (vsn) method with R package *vsn* (Huber *et al*, 2002). Missing values were imputed with *impute* function in R package *MSnbase* (Gatto & Lilley, 2012).

Since the principle of the SILAC labeling used in the experiment is to distinguish between newly synthesized and pre-existing proteins it also means that protein intensities of pre-existing proteins are expected to remain constant throughout the experiment; protein half-lives are generally much longer than the five hours used in the experiment (Mathieson *et al*, 2018, Schwanhauser *et al*, 2011). At the same time, the intensities of newly synthesized proteins are expected to increase during the experiment. Therefore, the normalization for the SILAC data was conducted with the normalization coefficients calculated for light data. In other words, the normalization was first done for the light data and the coefficients were then applied to heavy data. This approach maintains the main trends in the data for each SILAC fraction.

For the experiments with heat shock and recovery, fold changes between heat shock and mock shock were calculated. For heavy data, a fold change against pre-shocked

control was calculated to follow the accumulation of protein intensity of the newly synthesized proteins after mock or heat shock.

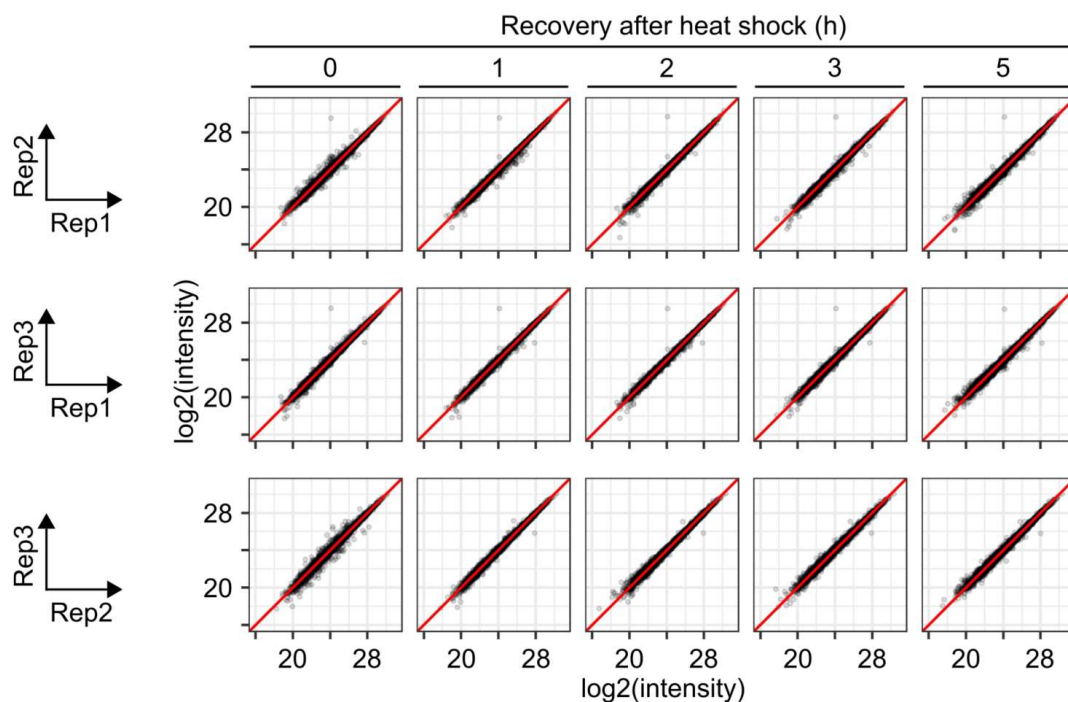
Heat shock-induced protein aggregation was estimated from heat shock to mock shock ratios right after the heat treatment from the pre-existing (light SILAC) fraction. A protein was assigned to be an aggregator if it had a significant and lower than  $\log_2(2/3)$  heat shock to mock shock ratio. The significance was estimated with limma analysis. A Benjamini-Hochberg adjusted p-value lower than 0.05 was considered significant.

For 2D-TPP data, the data was analyzed following previously described procedure (Becher *et al*, 2018). Within each condition (i.e. heat or mock shock and each replicate),  $\log_2$ -transformed and batch cleaned protein intensities from each temperature were normalized with *vs**n*. A ratio against 37°C was calculated for each temperature.

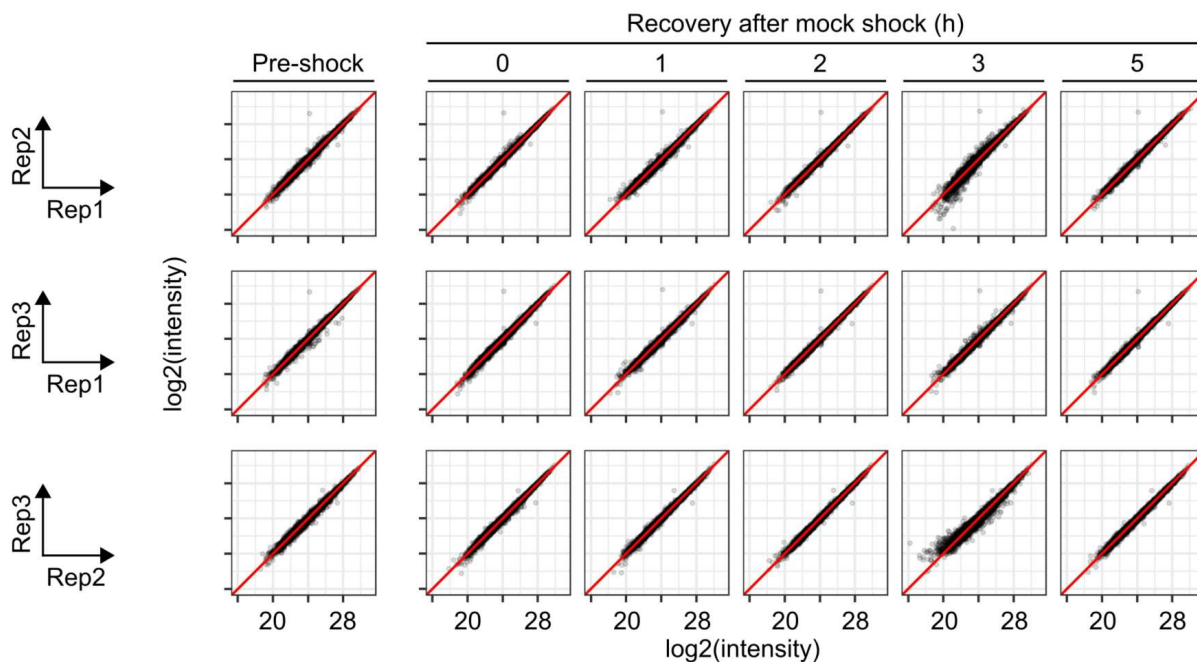
Change in thermal stability captured by 2D-TPP was estimated by calculating scores for thermal stability. To do this, the difference between heat shock and mock shock was calculated in each temperature; each data point was randomly chosen from one of the replicates. In a bootstrap loop, the process was iterated 500 times. Within each iteration, the differences were summed from every temperature. To remove the effect of heat-induced protein aggregation, the average difference from the first to temperatures was removed from all other temperature points. The aggregation adjusted sums from the 500 iterations were then transformed to z scores and present local stability scores. The mean of the scores was then tested for deviation from zero. The Benjamini-Hochberg adjusted p-value from that comparison presents local false discovery rate (FDR). For global stability score calculation, the mean local stability score of each protein was transformed to z scores. These global stability scores are reported in Figure 31-35.

In all experiments, the quantified protein intensities had a good correlation between replicates (Figure 1-5).

**A**



**B**



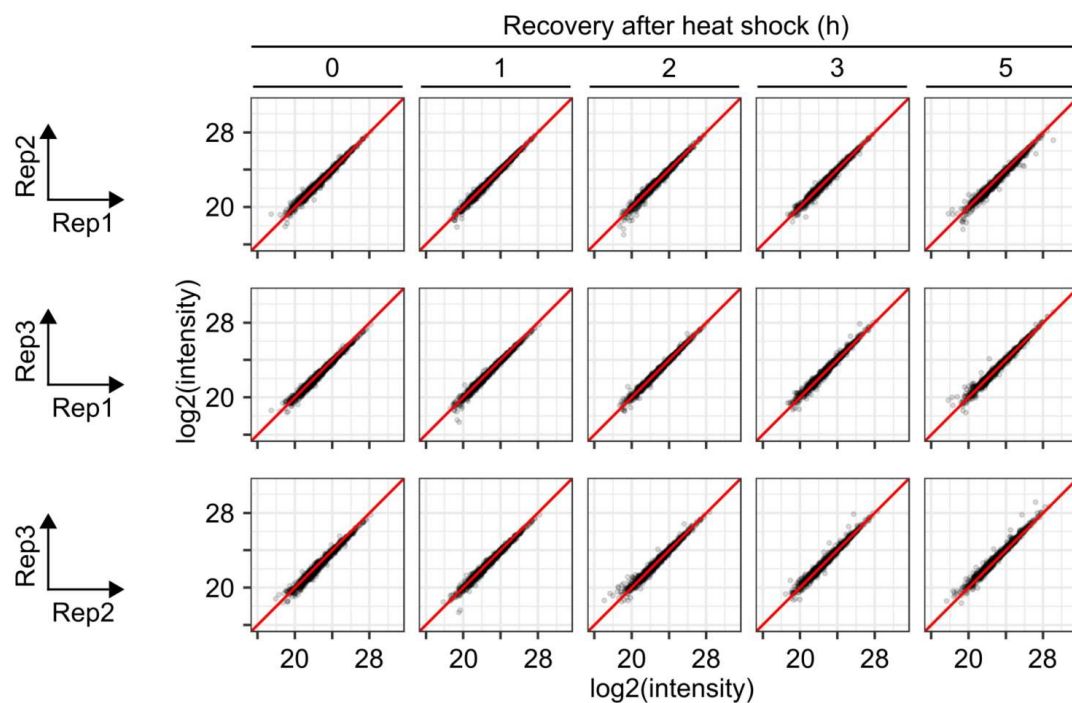
**Figure 1. Reproducibility between replicates in light SILAC dataset.**

A: Data from dynamic SILAC experiment with heat shock and recovery.

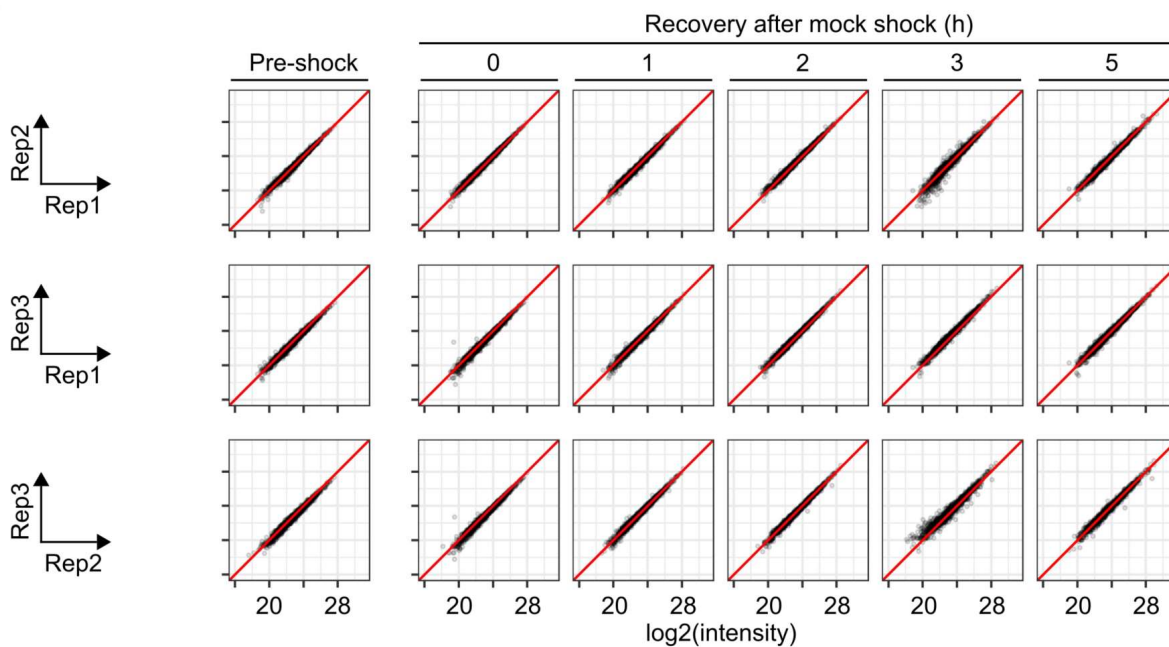
B: Data from dynamic SILAC experiment with mock shock and recovery

Proteins quantified from soluble fraction (cells lysed with mild detergent). Scatterplots showing normalized protein intensities in light SILAC fraction (pre-existing proteins).

**A**



**B**



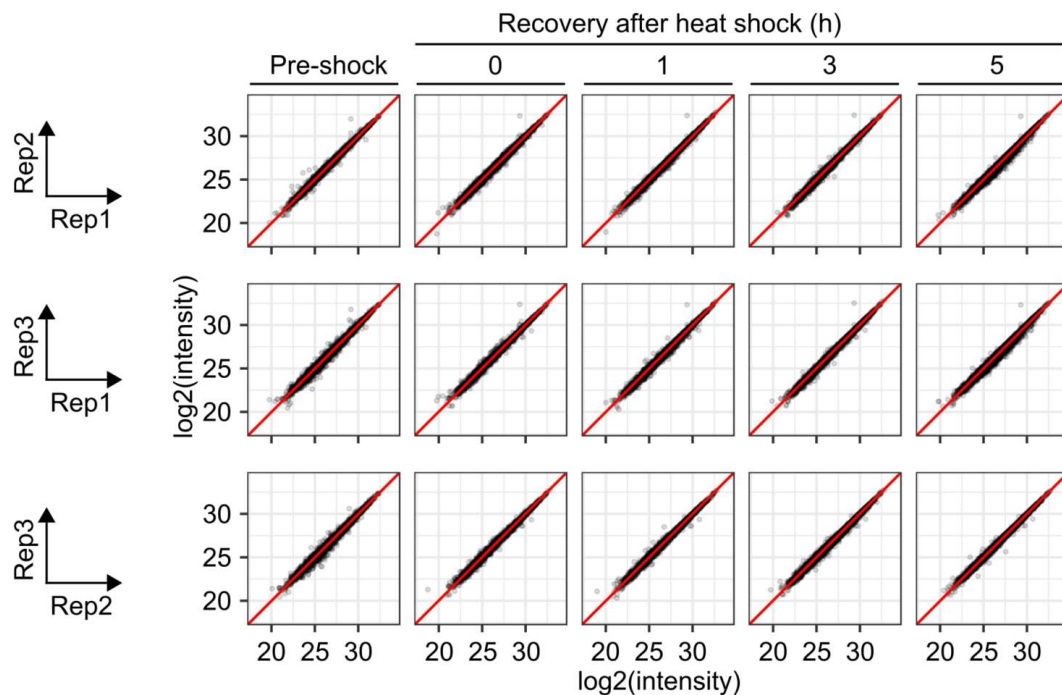
**Figure 2. Reproducibility between replicates in heavy SILAC dataset.**

A: Data from dynamic SILAC experiment with heat shock and recovery.

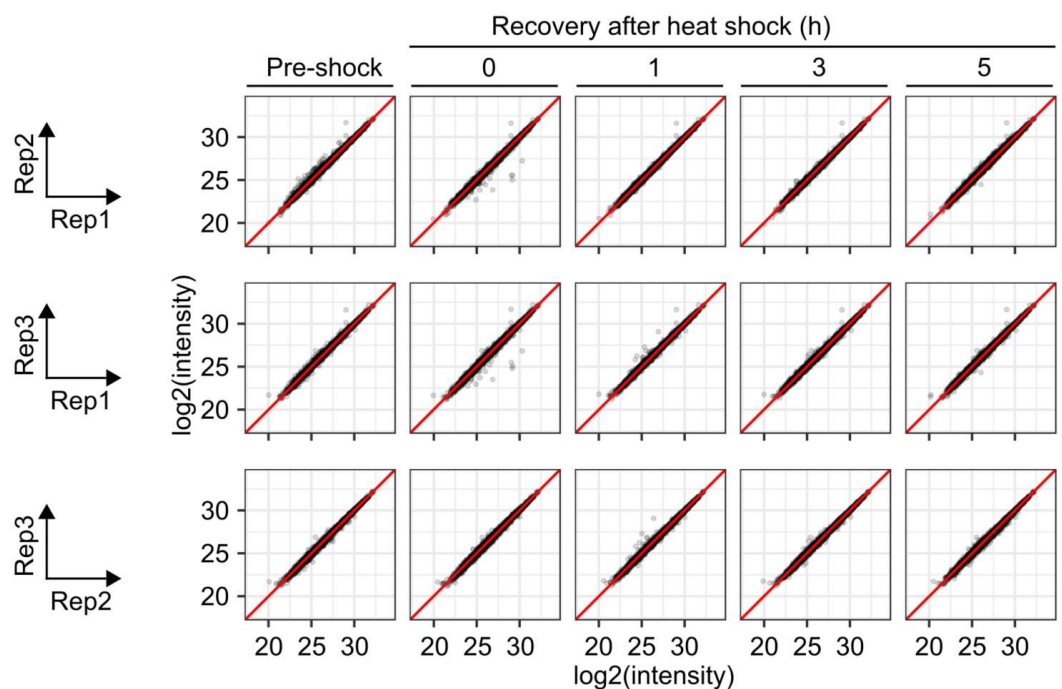
B: Data from dynamic SILAC experiment with mock shock and recovery

Proteins quantified from soluble fraction (cells lysed with mild detergent). Scatterplots showing normalized protein intensities in heavy SILAC fraction (newly synthesized proteins).

**A**



**B**

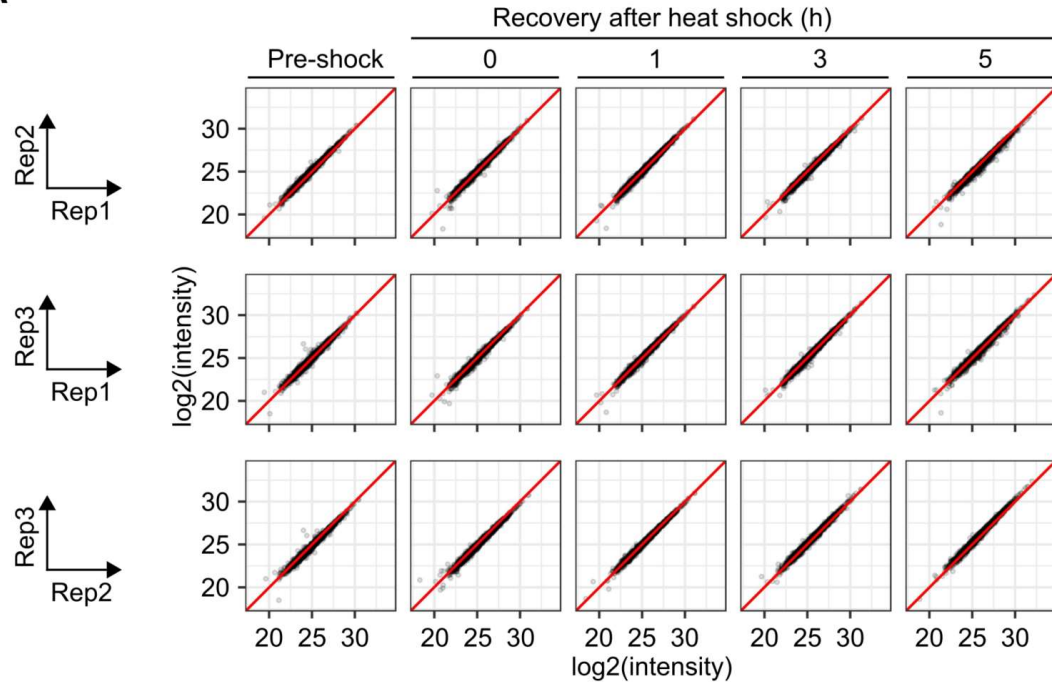
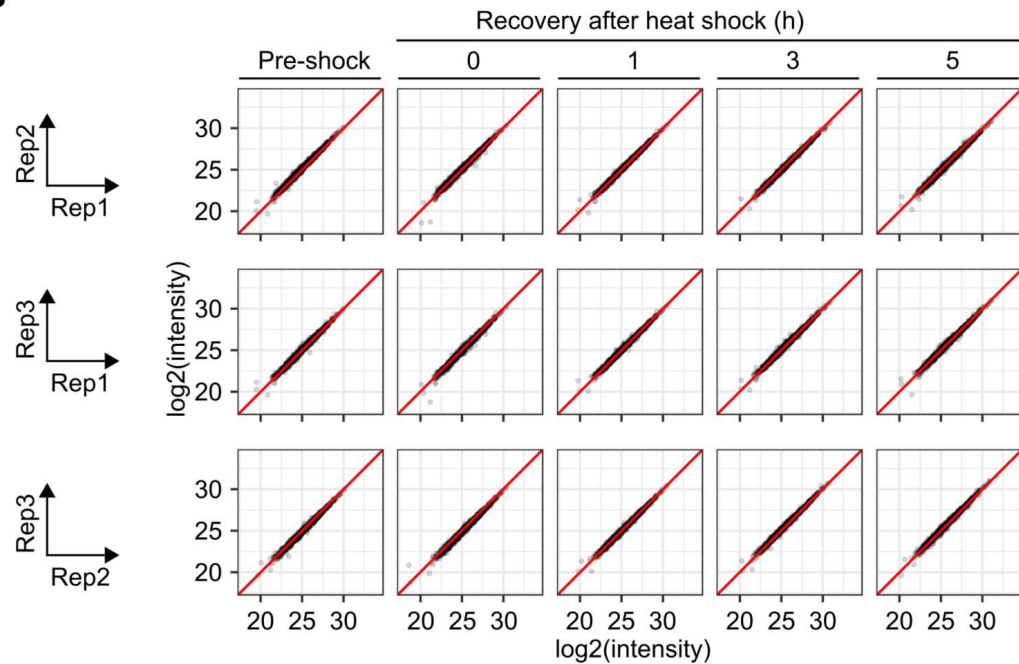


**Figure 3. Reproducibility between replicates in light SILAC dataset. Data from dynamic SILAC experiment with heat shock.**

A: Proteins quantified from soluble fraction (cells lysed with mild nonionic detergent; NP-40)

B: Proteins quantified from samples estimating the total protein amount (cells lysed with strong ionic detergent; SDS).

Scatterplots showing normalized protein intensities in light SILAC fraction (pre-existing proteins).

**A****B**

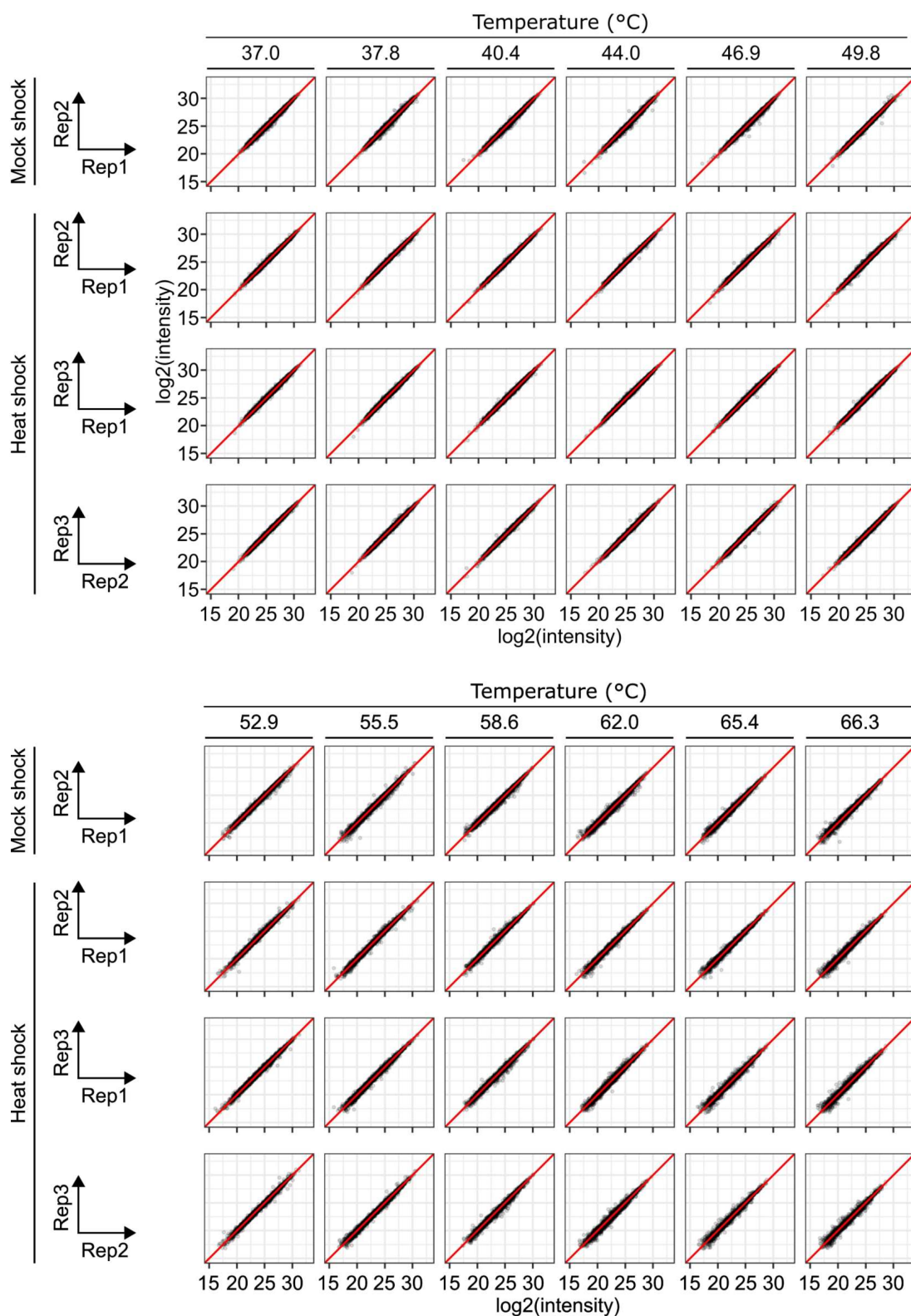
**Figure 4. Reproducibility between replicates in heavy SILAC dataset. Data from dynamic SILAC experiment with heat shock.**

A: Proteins quantified from soluble fraction (cells lysed with mild nonionic detergent; NP-40)

B: Proteins quantified from samples estimating the total protein amount (cells lysed with strong ionic detergent; SDS).

Scatterplots showing normalized protein intensities in heavy SILAC fraction (newly synthesized proteins).





**Figure 5. Reproducibility between replicates in two dimensional thermal proteome profiling.**

Scatterplots showing normalized protein intensities from soluble fraction in each temperature.



## 17. Protein features

For all protein features that were deduced from protein sequence, the amino acid sequence of the canonical isoform from UniProt (Consortium, 2019) was used. These features include gravity score, isoelectric point, molecular weight and secondary structure predictions.

Protein hydrophobicity was estimated by summing up gravity scores for each amino acid in the protein sequence. The gravity score for each amino acid is as follows: arginine (-4.5), lysine (-3.9), asparagine (-3.5), aspartate (-3.5), glutamine (-3.5), glutamate (-3.5), histidine (-3.2), proline (-1.6), tyrosine (-1.3), tryptophan (-0.9), serine (-0.8), threonine (-0.7), glycine (-0.4), alanine (1.8), methionine (1.9), cysteine (2.5), phenylalanine (2.8), leucine (3.8), valine (4.2) and isoleucine (4.5). Higher gravity score corresponds to higher hydrophobicity.

Isoelectric points and molecular weights were calculated with R package *peptides* (Osorio, 2015) using functions *pl* and *mw*, respectively.

Secondary structure elements were predicted using function *PredictHEC* from R package DECIPHER (Wright, 2016). The amount of each secondary structure (helix, sheet or random coil) in a protein sequence was calculated.

For all functions from the R packages mentioned above, the default settings were used.

# RESULTS

## 1. Experimental design

A workflow to measure protein solubility after heat shock and multiple time points during recovery was established. The workflow is based on hyperplexed MS that combines two different protein or peptide labeling methods. It allows for robust protein quantification from multiple samples simultaneously. Here, lysine and arginine amino acids in proteins were labelled with dynamic SILAC (dSILAC) (Doherty *et al*, 2009) and peptides from tryptic digestion were labelled with TMT tags.

dSILAC was used to differentiate newly synthesized proteins from pre-existing ones. This allowed to measure the solubility of mature proteins present in the cells during heat shock, as well as monitor the accumulation of proteins synthesized after heat shock. TMT labelling was used for pooling multiple samples and for protein quantification. These included samples collected at different time points after heat shock.

### 1.1 Dynamic stable isotope labeling by amino acids in cell culture

K562 human leukemia cells were grown at 37°C in light SILAC medium. At this stage, all proteins have adopted the light arginine and lysine amino acids. Before heat treatment, the medium was switched to heavy SILAC. After the medium switch, all newly synthesized proteins would adopt heavy lysine and arginine. At the same time, all mature proteins synthesized before medium switch contained light amino acids. To ensure complete consumption of light medium remaining in the cells, the medium switch was conducted 90 minutes before heat treatment. Without the lag time between medium switch and heat treatment, cells would still have residual light SILAC medium and newly synthesized proteins would incorporate light amino acids as long as they would be consumed. In this case, the changes in solubility of proteins labelled with light amino

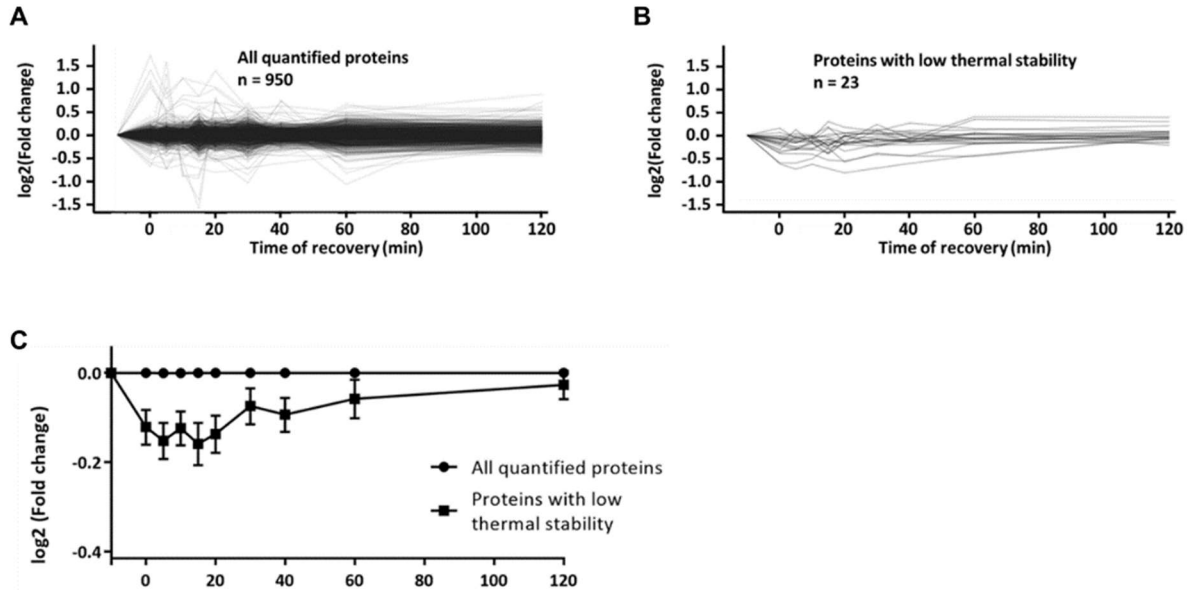
acids would be a combination of actual change in solubility and protein synthesis. Allowing cells to consume the residual light SILAC medium decreases this type of error. However, it should be noted that some light amino acids could be used in the protein synthesis if degradation of mature proteins and recycling of their amino acids would happen in the time scales used in the experiments.

## **1.2 Pilot experiment to measure aggregation and disaggregation**

A pilot experiment was conducted to test aggregation and disaggregation measurements with quantitative MS. Cells were exposed to a heat shock at 45°C for ten minutes in a temperature-controlled water bath. After heat shock, cells were let to recover at 37°C. Samples were collected 0, 5, 10, 15, 20, 30, 40, 60 and 120 minutes after heat shock. In addition, a control sample was collected before the heat shock.

After cell lysis, soluble protein fractions were collected, proteins digested to peptides and labelled with TMT. After pooling, samples were analyzed by quantitative mass spectrometry.

When compared to control (i.e. without heat shock), protein solubility stayed relatively constant after heat shock and during all time points of recovery (Figure 6A). In the MS analysis, only few proteins could be observed to show signs of aggregation or disaggregation (Figure 6A). However, when the data was analyzed separately for proteins with known low thermal stability in K562 cells (Savitski *et al*, 2014), it was noticed that they did show a trend of aggregation (Figure 6B-C). In addition, they showed a trend of disaggregation at the later time points; at the first 20 minutes the disaggregation was not apparent. However, the aggregation signal was too weak to be used in a proteome-wide MS analysis.



**Figure 6. Protein aggregation and disaggregation examined by quantitative mass spectrometry.**

Line graphs showing protein intensity in soluble fraction after heat shock and during recovery. Intensity values are compared to a non-heat shocked control sample (log<sub>2</sub>-transformed).

A All proteins

B Proteins with low thermal stability

C Summarized data for A and B

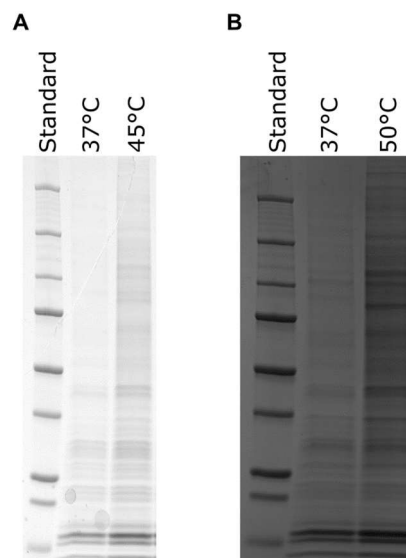
Data for proteins from the pre-existing (light) fraction.

### 1.3 Protein aggregation in K562 cells

The heat shock protocol was updated to test whether protein aggregation could be detected in MS analysis by using a different method for the heat shock. Instead of large cell volume (ml-scale) in a water bath, a heat block was used. It would also allow for smaller sample volumes (μl-scale) since samples could be placed on a 96-well plate. The heat block would also provide better control for the temperature. In addition, it was estimated that the reproducibility would increase since the heat shock in water bath required manual handling of the sample during the heat treatment.

The potential larger scale protein aggregation in K562 cells upon non-lethal heat shock was estimated by exposing cells to a heat shock and then analyzing the formation of protein aggregates with SDS-PAGE. Cells were exposed to either 37°C (negative control), 45°C or 50°C (positive control) for ten minutes. The heat shock was conducted in a temperature-controlled water bath. After cell lysis the insoluble protein fraction was collected and separated with SDS-PAGE, stained and analyzed.

In the control sample, some protein bands were visible in the insoluble fraction ('37°C' lanes in Figure 7) indicating the presence of some insoluble proteins. In contrast, after heat shock at 45°C the protein intensity in the insoluble fraction increased (Figure 7A). The increase was observed on wide spectrum of molecular weights. The same was observed for the positive control treated with 50°C heat shock (Figure 7B). The high molecular weight proteins appeared to increase intensity more in the aggregates after heat shock at 50°C when compared to heat shock at 45°C. Therefore, larger scale aggregation of proteins with ranging molecular weights was observed in non-lethal heat shock.



**Figure 7. Protein aggregation in K562 cells examined by SDS-PAGE.**

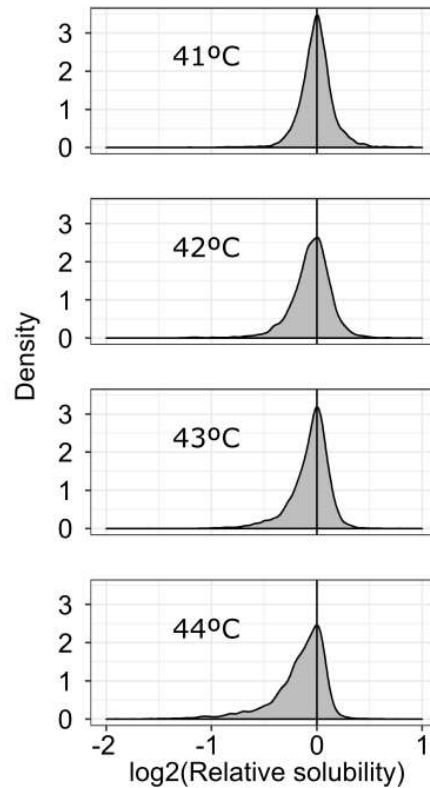
SDS-PAGE gels showing samples from treatment-control pairs. Insoluble protein fraction of samples after heat shock. For each treatment-control pair, same starting amount of cells were used.

A Heat shock at 45°C

B Heat shock at 50°C

## 1.4 Aggregation in updated heat shock method

To test whether the updated heat shock protocol would result in aggregation detectable by MS, cells were exposed to 37 (negative control), 41, 42, 43 and 44°C for ten minutes. The soluble protein fractions were collected from samples after heat shock and analyzed with MS. While the control-compared solubility was centered around zero (i.e. no change in log<sub>2</sub>-scale) at lowest heat shock temperatures in a bell-shaped distribution, the distribution started to have a left-skew when the heat temperature increased (Figure 8). Especially, at 44°C the left tail of the distribution was significantly increased indicating wide-spread protein aggregation.



**Figure 8. Protein aggregation at different heat shock temperatures examined by quantitative mass spectrometry.**

Proteins were exposed to indicated temperatures for ten minutes. The protein intensity in the soluble fraction was compared to a non-heat shocked control (log<sub>2</sub>-transformed).

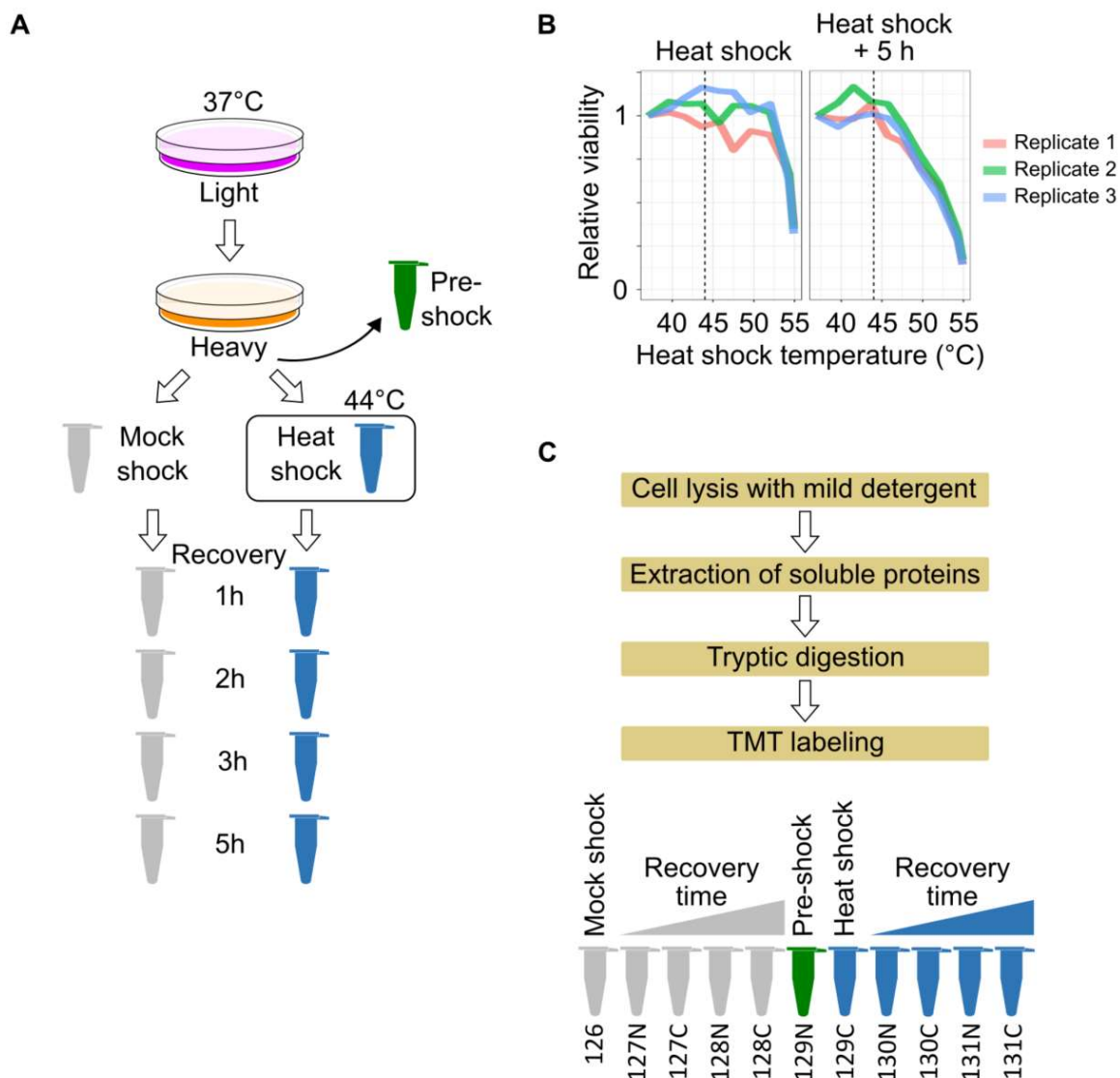
## **1.5 Finalized design: heat treatment, sample collection and quantitative mass spectrometry**

After the initial pilot experiments, the experimental design was finalized by changing the heat treatment from water bath to heat blocks, extending the time interval in sample collection during recovery and including time-matched control for each heat shocked sample.

Cells were distributed to two aliquots that were exposed either to a heat shock (ten minutes at 44°C) or mock shock (ten minutes at 37°C) (Figure 9A). A control sample was collected before the heat treatment (Figure 9A).

The length of the heat shock was chosen to be long enough to induce effects on the proteome but short enough to allow measurement of the instant effects before gene regulatory effects would emerge on protein level; extensive rewiring of transcription has been shown to take place after heat shock (Vihervaara *et al*, 2017). The length of the heat shock was also based on the report that K562 cells can tolerate a heat shock of ten minutes at 45°C (Schamhart *et al*, 1984). This tolerance was also tested and confirmed experimentally (as discussed next).

The heat shock temperature was chosen to be as high as possible (to gain a strong response) without compromising cell viability. The viability was an important aspect of the study since the recovery from the heat shock is assumed to be an active process requiring viable cells. To test this, cell viability after heat shock with different temperatures and after five hours of recovery was analyzed. It was found that K562 cells could tolerate a heat shock for ten minutes of temperatures up to approximately 53°C without any decrease in viability (Figure 9B). However, when analyzing the viability again after five hours of recovery at 37°C, the samples with heat shock at higher than approximately 45°C started to lose viability (Figure 9B). Therefore, heat shock at 45°C was the highest temperature that did not compromise cell



**Figure 9. Experimental design to study protein solubility after heat shock and during recovery.**

A: Workflow for dynamic SILAC experiment. K562 human cells were grown at 37°C in light SILAC medium. The medium was switched to heavy SILAC medium 90 minutes before heat treatment. The medium switch allowed to distinguish between mature pre-existing proteins (light) and newly synthesized proteins (heavy). For the heat treatment, cells were aliquot to be treated with either heat shock (ten minutes at 44°C) or a control mock shock (ten minutes at 37°C). Cells were let to recover at 37°C with samples collected after zero, one, two, three or five hours of recovery. In addition, a control sample was collected before the heat treatment.

B: Effects of heat shock to cell viability. Cell viability as a function of heat shock temperature as measured immediately after the heat shock (left panel) or after five hours of recovery (right panel). Vertical lines indicate 44°C used in the experiment described in A.

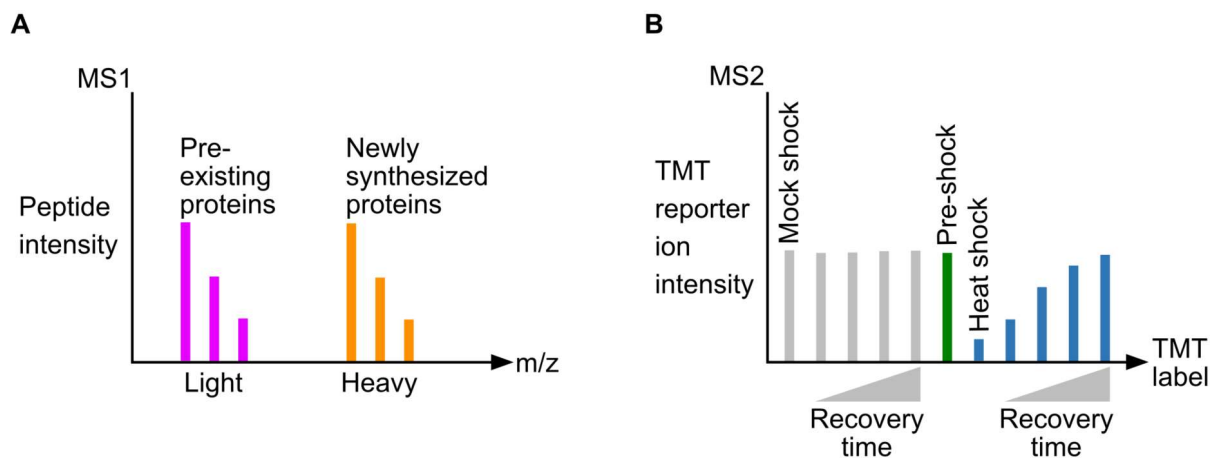
C: Sample processing for multiplexed LC-MSMS analysis. Cells were lysed using a mild non-ionic detergent (NP-40) allowing the extraction of the soluble proteome. Proteins were digested with trypsin to gain peptides compatible for bottom-up MS analysis. Peptides from each sample were labelled with different TMT tags that allows pooling of the samples to a one MS experiment.



viability in the time scale of the experiment. For a safety measure, one degree lower, i.e. 44°C, was chosen for the heat shock temperature in the assay.

After heat treatment, cells were let to recover at 37°C for zero, one, two, three or five hours (Figure 9A). After sample collection, cells were lysed with non-ionic detergent (NP-40) that does not solubilize protein aggregates (Reinhard *et al*, 2015). Soluble protein fraction was gathered and digested with trypsin to peptides compatible for mass spectrometry analysis. Peptides from eleven different samples were labelled with TMT, pooled to a one mass spectrometry sample and analyzed with an orbitrap mass spectrometer (Figure 9C).

In the MS analysis, peptides originating from pre-existing (light) and newly synthesized fraction (heavy) can be distinguished in the MS1 scan (Figure 10A). This allows to separately analyze the two SILAC fractions. In the following MS2 scan, peptides from each experimental condition are identified and quantified (Figure 10B). The relative quantification (based on TMT reporter ion intensities) of proteins from the soluble fraction is later used as a measure for solubility.



**Figure 10. Mass spectrometry analysis of dynamic SILAC data.**

A: Separation of SILAC fractions during MS1 analysis. Peptides from the SILAC-labelled pre-existing (light) and newly synthesized (heavy) proteins can be distinguished by their mass difference in MS1 analysis and, therefore, analyzed separately.

B: Peptide quantification during MS2 analysis. The peptides from different samples can be distinguished and quantified based on the TMT reported ion intensity. TMT reporter ions with different masses are resulting from fractionation of the parent TMT-tags prior to MS2 analysis. The bar plot shows an example of a peptide originating from an aggregating and disaggregating protein.

To estimate changes in the total protein amount, the experiment was repeated using additionally a lysis protocol (with strong ionic detergent - SDS) that would bring all proteins to the solution. In this setup, samples were collected before heat shock and after zero, one, three and five hours after heat shock.

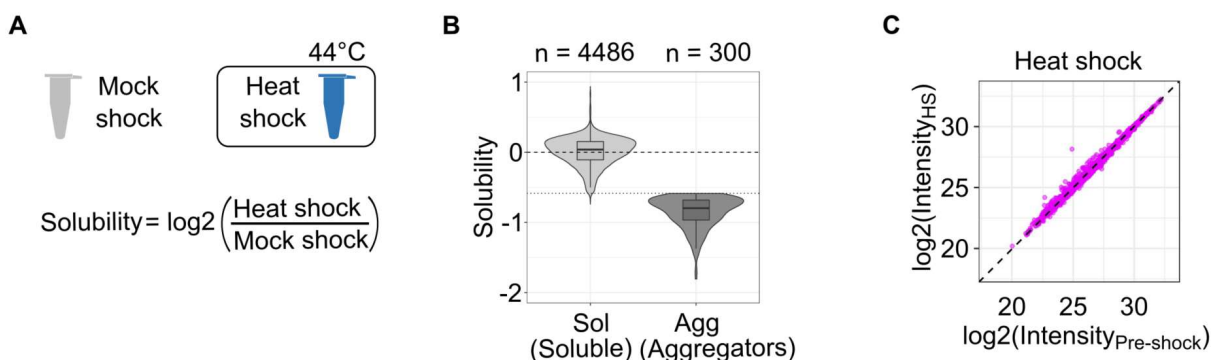
The dataset (soluble fraction) included quantification of 7226 proteins. To select only for high quality data, two quality control filters were used. First, all proteins that were quantified with only one unique peptide were omitted. Secondly, all proteins that were detected only from one replicate were omitted. The second step was conducted separately for light and heavy SILAC channels. After the quality control steps, the high quality dataset included 4786 proteins from the pre-existing protein fraction (light) and 1296 proteins from the newly synthesized protein fraction (heavy). The samples estimating the total protein amount resulted in a high quality dataset that included 5548 proteins from the pre-existing protein fraction (light) and 1455 proteins from the newly synthesized protein fraction.

## **2. Proteins aggregating in transient heat shock**

Protein solubility in pre-existing fraction (light) after heat shock was estimated by calculating a ratio between heat and mock shock sample from protein intensities in the soluble fraction (Figure 11A). The aggregating proteins (from now on 'aggregators') were defined as proteins for which the heat shock to mock shock ratio was lower than  $\log_2(2/3)$  (Figure 11B). In addition, the ratios were tested for statistical significance. For protein to be an aggregator, it was also required that the p-value in limma analysis (adjusted for multiple hypothesis testing) was below 0.05.

Using the criteria described above, 300 proteins were found to aggregate in the heat shock while 4486 proteins did not show change in solubility after heat shock (Figure 11B). The aggregators comprise approximately 6% of the quantified proteins. The

largest decreases in solubility were approximately 70% (corresponding to -1.75 on log2-scale).



**Figure 11. Solubility change after heat shock.**

A: Solubility was measured as log2-transformed ratios of protein intensity in the soluble fraction between heat and mock shocked samples collected right after the heat treatment.

B: Definition of aggregating proteins. Aggregators were defined as proteins having a solubility lower than  $\log_2(2/3)$  with Benjamini-Hochberg adjusted p-value lower than 0.05. All other proteins were defined as soluble. The horizontal dotted line shows the cut-off value at  $\log_2(2/3)$ .

C: Total protein levels in heat shocked sample and pre-shock control. Total protein intensities from samples lysed with strong ionic detergent (SDS). Scatterplot comparing intensities of heat shocked and pre-shock control sample.

All data shown for pre-existing protein fraction (light).

The five most aggregating proteins included the most aggregating protein lymphoid-specific helicase (HELLS; solubility = -1.81), followed by sex comb on midleg-like protein 2 (SCML2; -1.72), nucleolar MIF4G domain-containing protein 1 (NOM1; -1.67), heterogeneous nuclear ribonucleoprotein H2 (HNRNPH2; -1.50) and nucleolar protein 9 (NOP9; -1.47).

Using the same criteria as for aggregators but with changing the sign of the fold change cut-off, five proteins were found to have increase in solubility upon heat shock. These proteins were probable methyltransferase TARBP1 (solubility = 0.79), basement membrane-specific heparan sulfate proteoglycan core protein HSPG2 (0.67), serine/threonine-protein kinase LMTK2 (0.59), transforming growth factor-beta-induced protein (TGFB1; 0.85) and C-type mannose receptor 2 (MRC2; 0.83). These proteins might reflect false positives, since only five of them were found and they do not have an

insoluble sub-population (a concept described later in section 2.3) from where the increase of solubility could stem from.

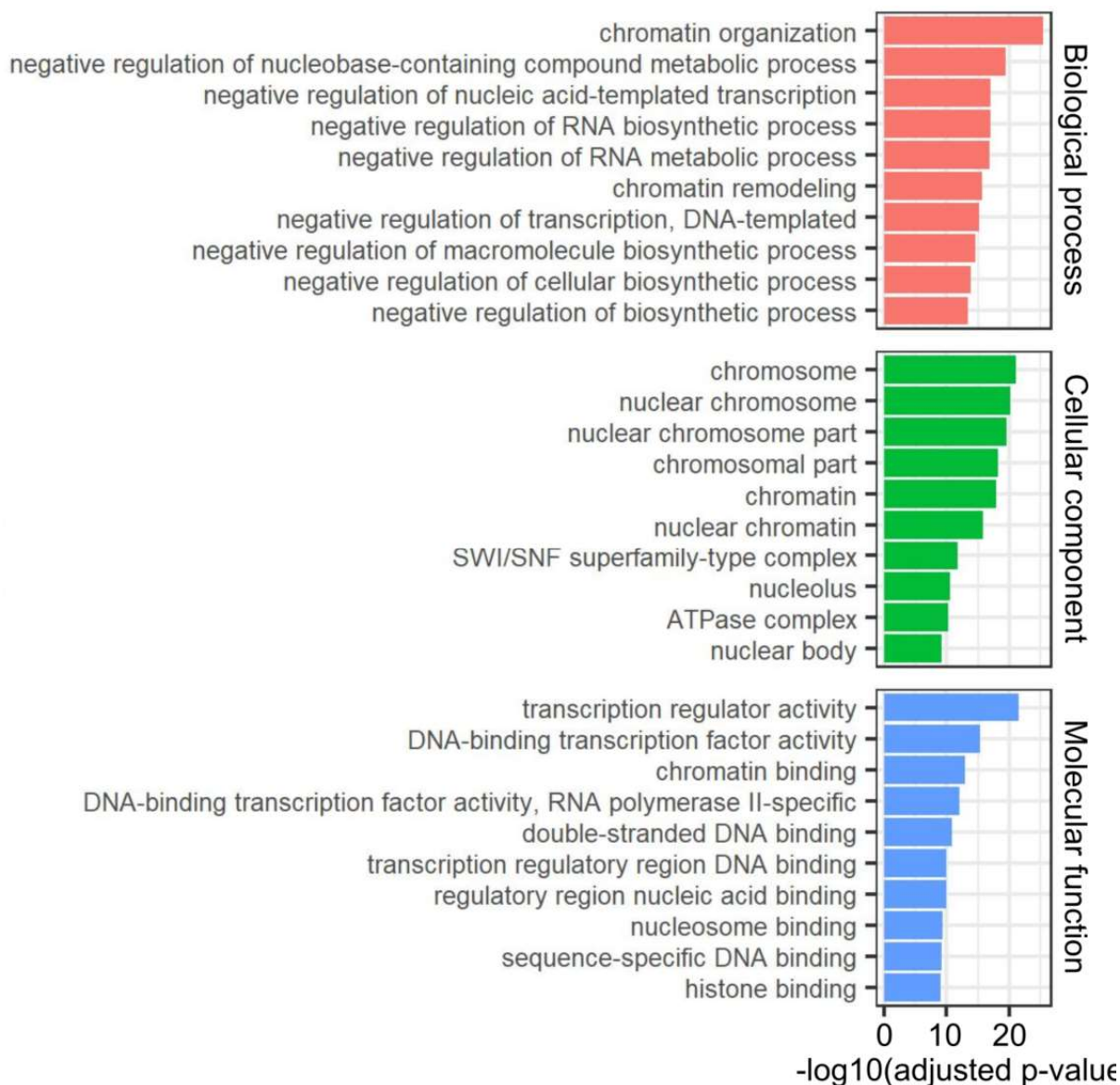
The loss of protein intensity in the soluble fraction after heat shock could result from protein degradation induced by the heat shock. Therefore, the total protein amount in the samples were estimated from samples lysed with strong ionic detergent (SDS). The total protein intensity remained unchanged between heat shock and mock shock (Figure 11C). This indicates that there was no immediate protein degradation, or any other change in the total protein amount, after heat shock. Therefore, the loss of intensity in the soluble fraction most probably stems from formation of insoluble aggregates.

Previous studies have shown that the protein aggregation can correspond to amorphous aggregates (Wang *et al*, 2010), fibers (Bauerlein *et al*, 2017, Knowles *et al*, 2014, Wang *et al*, 2010) or phase separated membrane-less organelles (Brangwynne *et al*, 2009, Riback *et al*, 2017), for example. The aggregates can also include co-aggregating macromolecules, such as RNA (Saad *et al*, 2017). It is important to note that the aggregation assay here is based on protein quantification from soluble fraction and includes information only about protein solubility, not the structural state of the aggregate. Therefore, the definition of aggregator here is quite general and might include multiple types of aggregates.

## **2.1 Aggregators are enriched in nuclear proteins**

The functionalities of aggregators were analyzed by performing enrichment analysis for gene ontology (GO) terms. Multiple terms within all GO domains were found to be enriched in aggregators with high statistical significance (Figure 12). The aggregators had an enrichment in GO terms related to nuclear localization and nuclear functions, such as DNA binding, chromatin organization and transcription regulator activity (Figure 12). The enriched terms also included protein complexes (SWI/SNF superfamily-type

complex and ATPase complex). It is note-worthy, that all enriched GO terms listed in Figure 12 are related to nucleus; although only the ten most enriched terms are shown.



**Figure 12. Functional characterization of aggregators.**

GO term enrichment analysis of aggregating proteins. Bar plot showing the top ten most significant (lowest p-values) GO terms enriched in aggregators in the three GO domains (biological process, cellular component and molecular function).

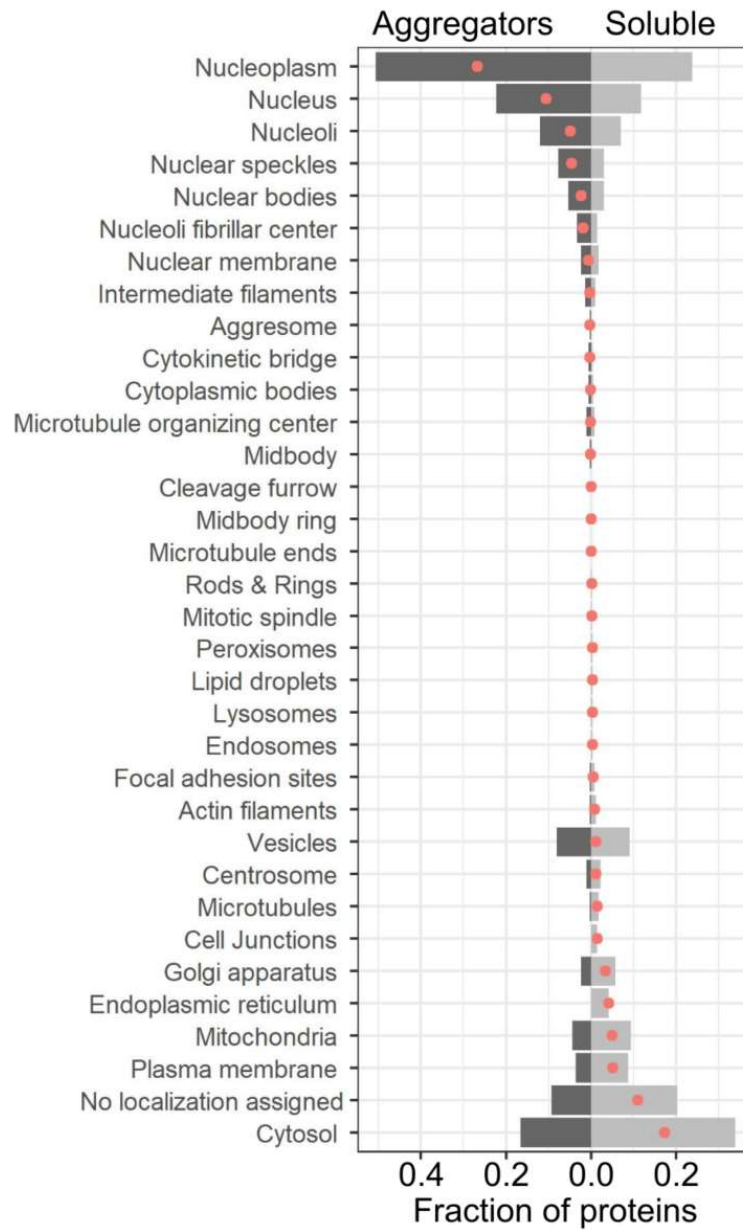
To gain a deeper view on localization of aggregators - and proteins that did not aggregate - localization annotation from Human Protein Atlas (Thul *et al*, 2017) were analyzed. The fraction of proteins containing a particular localization annotation was calculated (separately for aggregators and soluble proteins).

The most frequent annotations for aggregators were nucleoplasm, nucleolus, nucleoli and other nuclear compartments (Figure 13). In contrast, the most frequent annotation for soluble proteins was cytosol (Figure 13). When comparing aggregating and soluble proteins, all annotations related to nucleus had higher occurrence in aggregating proteins; the opposite was true for all major cytoplasmic compartments, i.e. soluble proteins had the highest fraction of proteins with the annotation plasma membrane, mitochondria, endoplasmic reticulum and Golgi apparatus (Figure 13).

## **2.2 Stress granule proteins remained soluble after heat shock**

One very well documented phenomenon related to heat and other stresses is the formation of stress granules (Collier *et al*, 1988, Collier & Schlesinger, 1986, Ivanov *et al*, 2018). Proteins related to stress granules have been reported to aggregate in yeast upon heat shock (Grousl *et al*, 2013, Wallace *et al*, 2015). To investigate stress granule formation, the solubility of proteins related to stress granules was analyzed. From the dataset, 44 proteins with a GO term 'cytoplasmic stress granule' was found.

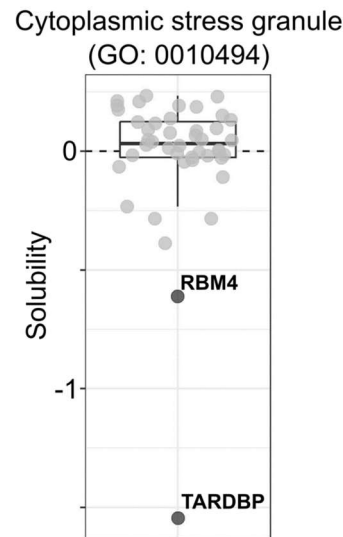
When analyzing the heat shock-induced solubility change, it was found that only two of the stress-granule proteins aggregated: TAR DNA-binding protein 43 (TARDBP) and RNA binding protein 4 (RBM4) (Figure 14). The solubility after heat shock for RBM4 was -0.70, which is quite close to the cut-off used for aggregators; few other proteins not classified as aggregators can also be seen to lose some solubility when compared to the majority of stress granule proteins in Figure 14. However, TARDBP with a solubility of -1.41 clearly has a different aggregation propensity than the other stress



**Figure 13. Localization of aggregators.**

Localization of aggregators and soluble proteins. Bar plot showing the fraction of aggregators (dark grey) and soluble proteins (light grey) with certain localization annotations. Red dots indicate the difference between aggregators and soluble proteins.

granule proteins. With these exceptions, the majority of stress granule proteins, however, remain soluble after heat shock.



**Figure 14. Solubility of stress granule proteins.**

Solubility after heat shock of proteins related to stress granules. Solubility values of proteins with a GO term 'cytoplasmic stress granule' (GO:0010494). Aggregators are highlighted with dark grey and annotated.

## 2.3 Insoluble sub-populations

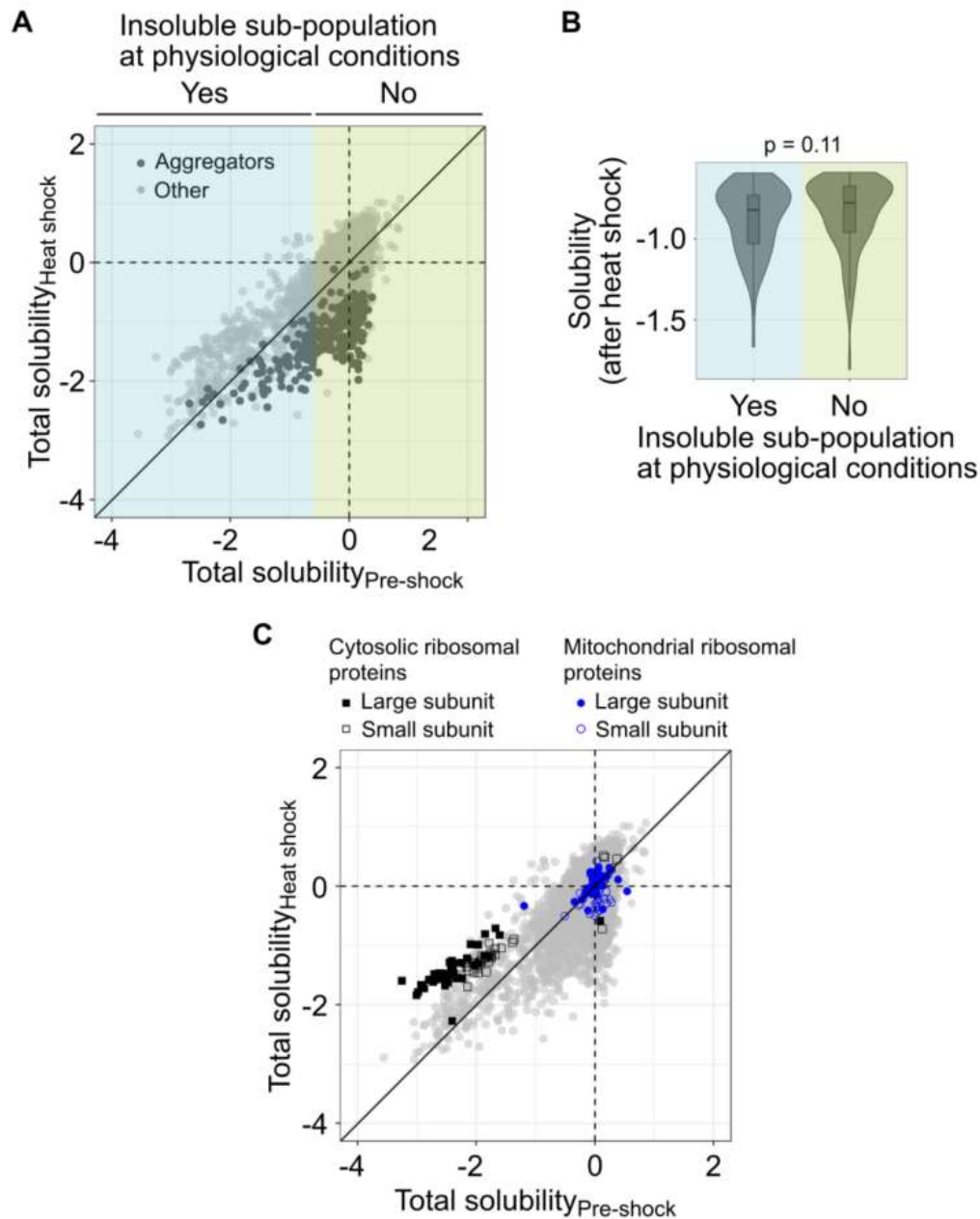
Protein aggregation has been linked to stress conditions - especially heat stress. However, it has been shown that proteins can contain a sub-population that is insoluble at normal physiological conditions (Becher *et al*, 2018, Ibstedt *et al*, 2014, Sridharan *et al*, 2019). To examine the presence of insoluble protein sub-populations, protein intensities in the soluble protein fraction were compared to total protein intensity. Similar intensities in both fractions indicate that protein is completely present in the soluble fraction. If protein intensity is lower in the soluble fraction, it indicates the presence of an insoluble sub-population.



It was found that, at physiological conditions, most proteins were soluble (Figure 15A); the total solubility values in the samples collected before heat shock (x-axis in Figure 15A) were densely centered around zero (i.e. no change in log<sub>2</sub>-scale). However, a fraction of proteins had an insoluble sub-population (Figure 15A). This can be seen as the right-skewed distribution of total solubility values on x-axis in Figure 15. The most insoluble proteins had approximately only 25% of protein mass in the soluble fraction (i.e. total solubility of -2 on the log<sub>2</sub>-scale), although in the extreme cases the total solubility was even below 10%.

The presence of insoluble sub-population could impact the aggregation-propensity of a protein. For example, the presence of an insoluble fraction could indicate that the protein is unstable. On the other hand, if the protein is already in the aggregates, the heat shock might not have further impact on its solubility. It was found that, after heat shock, protein aggregation was observed regardless of the presence of insoluble sub-population at physiological conditions: the aggregators can be found from proteins with different total solubility in the absence of stress (Figure 15A). Furthermore, when grouping proteins according to the presence of insoluble sub-population, the solubility change of aggregators was similar within the two groups (Figure 15B).

In yeast, it has been reported that ribosomal proteins are partly insoluble in unstressed conditions (Ibstedt *et al*, 2014, Weids *et al*, 2016). To examine whether the same holds true in human cells, the total solubility of ribosomal proteins was analyzed. It was found that cytosolic ribosomal proteins had an insoluble sub-population at physiological conditions (Figure 15C). On the contrary, mitochondrial ribosomal proteins were completely soluble (Figure 15C). In addition, there was a slight trend towards proteins from large subunit to be more soluble than proteins from the small subunit (Figure 15C). Upon heat shock, mitochondrial ribosomal proteins stay soluble while cytosolic ribosomal proteins have a slight trend of becoming more soluble (Figure 15).



**Figure 15. Insoluble protein sub-populations.**

A: Total protein solubility before and after heat shock. Total solubility was measured as a ratio between protein intensity in the soluble fraction and the total protein intensity. Scatterplot comparing total solubility from samples before and after heat shock. Aggregators are highlighted as dark grey. Proteins with a total solubility below -0.6 in the pre-shock sample were assigned as proteins with an insoluble sub-population.

B: Solubility after heat shock of proteins with or without an insoluble sub-population. Combined violin and box plots showing solubility values. P-value is shown for non-parametric Wilcoxon test.

C: Total protein solubility before and after heat shock. As in A, except the ribosomal proteins from small (empty symbols) and large (filled symbols) subunits are highlighted from cytosolic (black) and mitochondrial (blue) ribosomes.

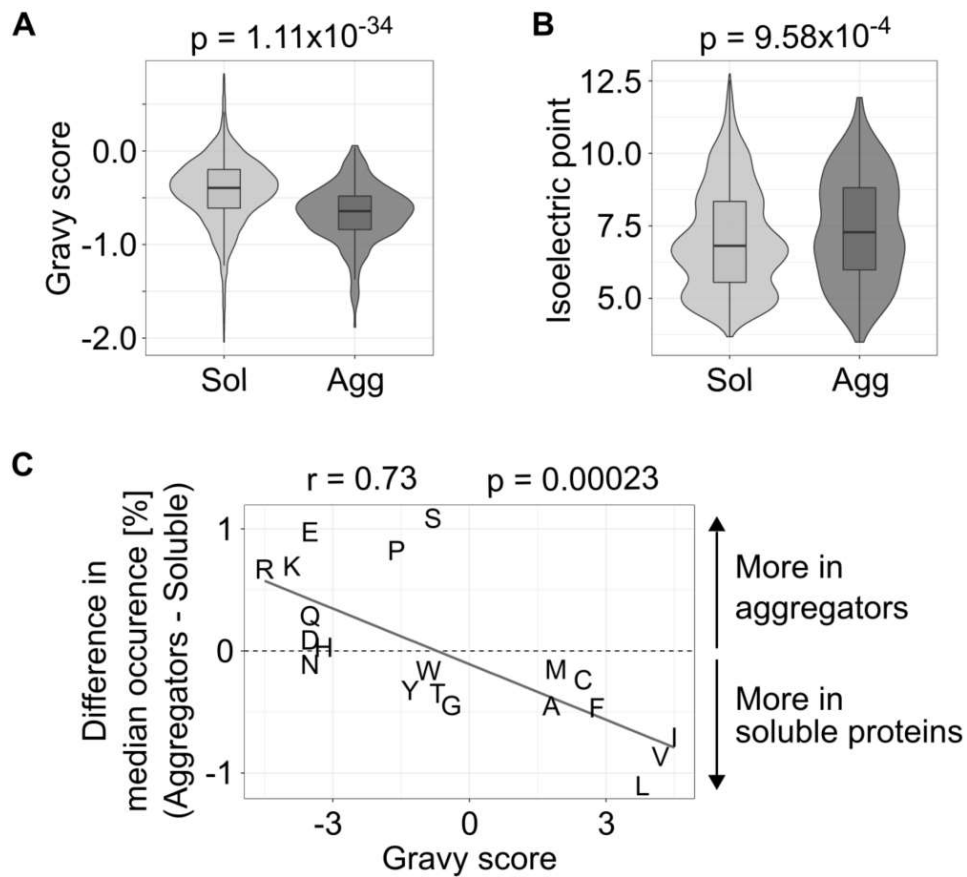
## 2.4 Physicochemical properties of aggregators

The aggregation of a protein in a mild heat shock could be determined by its physicochemical properties. To investigate this, selected physicochemical features were compared between aggregating and soluble proteins.

One feature that possibly effects the aggregation propensity of a protein is its hydrophobicity, especially considering the role of hydrophobic regions in protein folding. Hydrophobicity was estimated by calculating a gravity score for each protein. The values were then compared between aggregators and soluble proteins.

Aggregators were found to be more hydrophilic than soluble proteins (Figure 16A,  $p = 1.11 \times 10^{-34}$ ). The difference between the medians was approximately 0.25. To put it in context, the difference in gravity score for isomeric amino acids is 0.7 for leucine and isoleucine, for example.

A measure related to hydrophobicity is isoelectric point. It is the pH where a protein would have a zero net charge. The charge of a protein is due to residues that can be in a charged state. In neutral pH, glutamate and aspartate contribute to negative charge while arginine and lysine for positive charge. In addition, histidine can adopt charged sates (both positive and negative) in a biologically relevant pH range. It should be noted, that the charge state of these residues at particular pH is determined by their acid constant, which in turn is determined by the chemical environment. Therefore, for example, the charge state of histidine can change from positive to neutral, and to negative depending on the environment; a net positive charge near histidine will impact its acidic constant to favor negative charge and *vice versa*. Stronger acids like in the side chains of glutamine and aspartate, on the other hand, are less sensitive for the chemical environment.



**Figure 16. Physicochemical features of aggregators.**

A: Combined violin- and boxplots showing hydrophobicity (gravy score) of aggregators and soluble protein.

B: Combined violin- and boxplots showing isoelectric points of aggregators and soluble protein.

C: The difference in median amino acid occurrence for each amino acid between aggregators and soluble proteins as a function of hydrophobicity (gravy score). Pearson coefficient ( $r$ ) and  $p$ -value is shown for correlation analysis.

$P$ -values in A and B are for non-parametric Wilcoxon test.

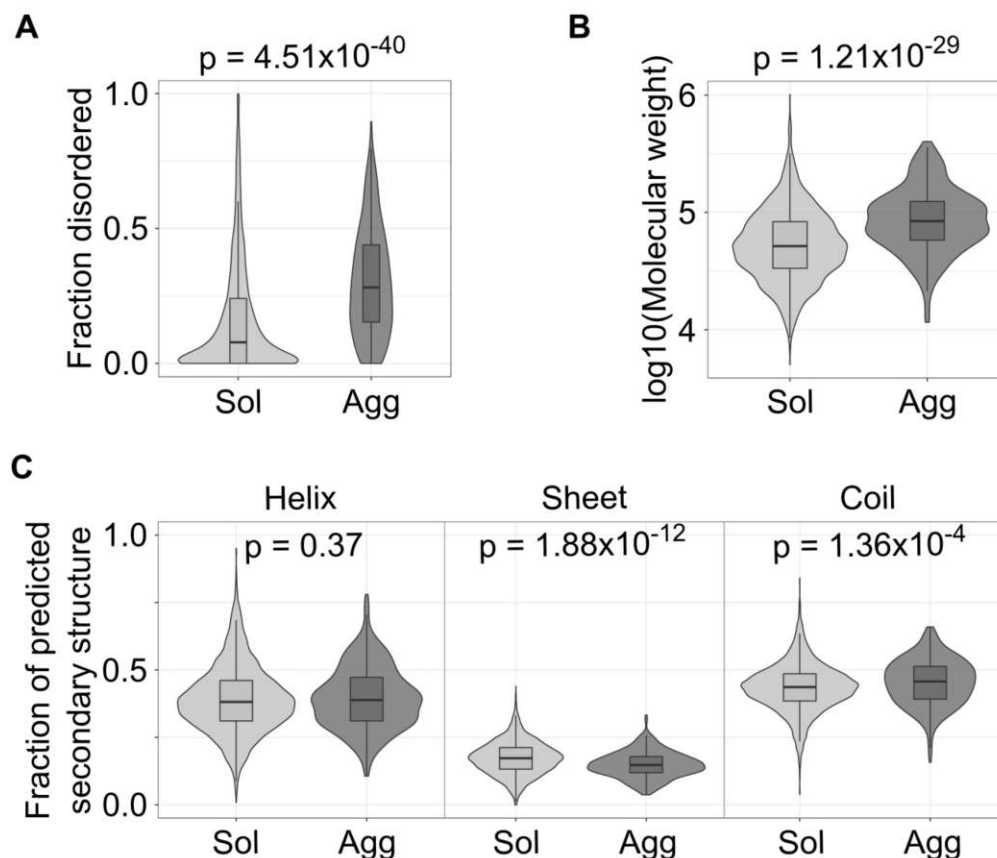
Aggregating proteins were found to have higher  $pI$  than the soluble proteins (Figure 16B,  $p = 9.58 \times 10^{-4}$ ), i.e. they are more positively charged. The difference between the medians was 0.47. In the context of heat shock, a drop in  $pH$  has been observed with some model organisms (Drummond *et al*, 1986, Srinivas & Revathi, 1993).

To have a deeper view on the enrichment of hydrophobicity and high isoelectric point in aggregators, an amino acid level analysis was performed. The occurrence of each amino acid within each protein was calculated. Then the occurrence was compared between aggregators and soluble proteins, and analyzed in the context of hydrophobicity.

When comparing gravity scores (hydrophobicity) of individual amino acids to their occurrence in aggregators and soluble protein, a significant negative correlation was observed (Figure 16C,  $p = 0.00023$ ). As discussed above on protein level, aggregators tend to be more hydrophilic. The amino acid level analysis further revealed that the observation was not due to any single amino acid but a clear trend was observed for enrichment of hydrophilic amino acids in aggregators, as well as depletion of hydrophobic amino acids. In addition, the higher isoelectric point in aggregators could be explained by the increased amounts of positively charged amino acids (lysine and arginine); depletion of negatively charged amino acids (glutamate and aspartate) was not observed (Figure 16C). The notable exceptions from the trend were proline and serine which had higher occurrence in aggregators. In contrast to serine, the two other amino acids with hydroxyl groups in the side chains, tyrosine and threonine, were slightly enriched in soluble proteins while serine was the most enriched amino acid in aggregators.

## **2.5 Structural features of aggregators**

Intrinsically disordered regions are stretches of primary structure in proteins that do not possess any defined secondary structure (Dyson & Wright, 2005). Their presence has been previously linked to protein aggregation (Hosp *et al*, 2017, Uemura *et al*, 2018, Walther *et al*, 2015). To analyze whether there is a link between disordered regions and aggregation, the fraction of protein sequence predicted to be disordered was acquired from a database and compared between aggregators and soluble proteins.



**Figure 17. Structural features of aggregators.**

A: Combined violin- and boxplots showing the fraction of intrinsically disordered regions in aggregators and soluble protein.

B: Combined violin- and boxplots showing the molecular weight of aggregators and soluble protein.

C: Combined violin- and boxplots showing the fraction of secondary structure elements in aggregators and soluble protein.

P-values are for non-parametric Wilcoxon test.

When analyzing the proportion of a protein sequence predicted to be intrinsically disordered, it was found that aggregating proteins contain higher amount of them when compared to soluble proteins (Figure 17A,  $p = 4.51 \times 10^{-40}$ ). The aggregators had a median of 28.2% of the protein sequence predicted as disordered while the median was

only 7.9% for soluble proteins. In addition, the shape of the distribution is quite different between the two. Soluble proteins have the highest density at the bottom of the values, i.e. most soluble proteins do not contain (or contain only a small fraction of) disordered regions. On the other hand, aggregators have the highest density at approximately 20% of disordered regions. In addition, only a small fraction of aggregators lacks disordered regions completely.

High molecular weight adds complexity to protein folding making proteins more prone for misfolding. In addition high molecular weight has been associated with aggregation propensity (Hosp *et al*, 2017, Kramer *et al*, 2012, Weids *et al*, 2016). Therefore, the molecular weights of aggregators and soluble proteins were analyzed.

The analysis showed that the molecular weight was significantly higher for aggregators (Figure 17B,  $p = 1.21 \times 10^{-29}$ ). The difference between the medians was approximately 1.74 Da (0.24 on the log10-scale). Although the difference corresponds to less than two hydrogen atoms, the trend is clear.

The strong enrichment of intrinsically disordered regions in aggregating proteins suggested to further explore the (lack of) defined secondary structure elements. The presence of defined secondary structure elements (alpha helices, beta sheets and random coil) was predicted for each protein and their occurrence was compared between aggregators and soluble proteins.

The content of alpha helices was similar between the two groups (Figure 17C,  $p = 0.37$ ). The median fraction of predicted alpha helix secondary structure in soluble proteins was 38.2% and for aggregators 38.9%. A significant enrichment in beta sheets was observed in soluble proteins (Figure 17C,  $p = 1.88 \times 10^{-12}$ ). If measured with the median of the predicted fraction, soluble proteins contained 17.3% beta sheets while aggregators had 14.8%. Therefore, aggregators were predicted to contained 14.5% less beta sheets. Similar to the results obtained for intrinsically disordered regions, a significant enrichment in random coil secondary structure was found in aggregating

proteins (Figure 17C,  $p = 1.36 \times 10^{-4}$ ). The median for soluble proteins was 43.7% of predicted random coil and for aggregators it was 45.7%.

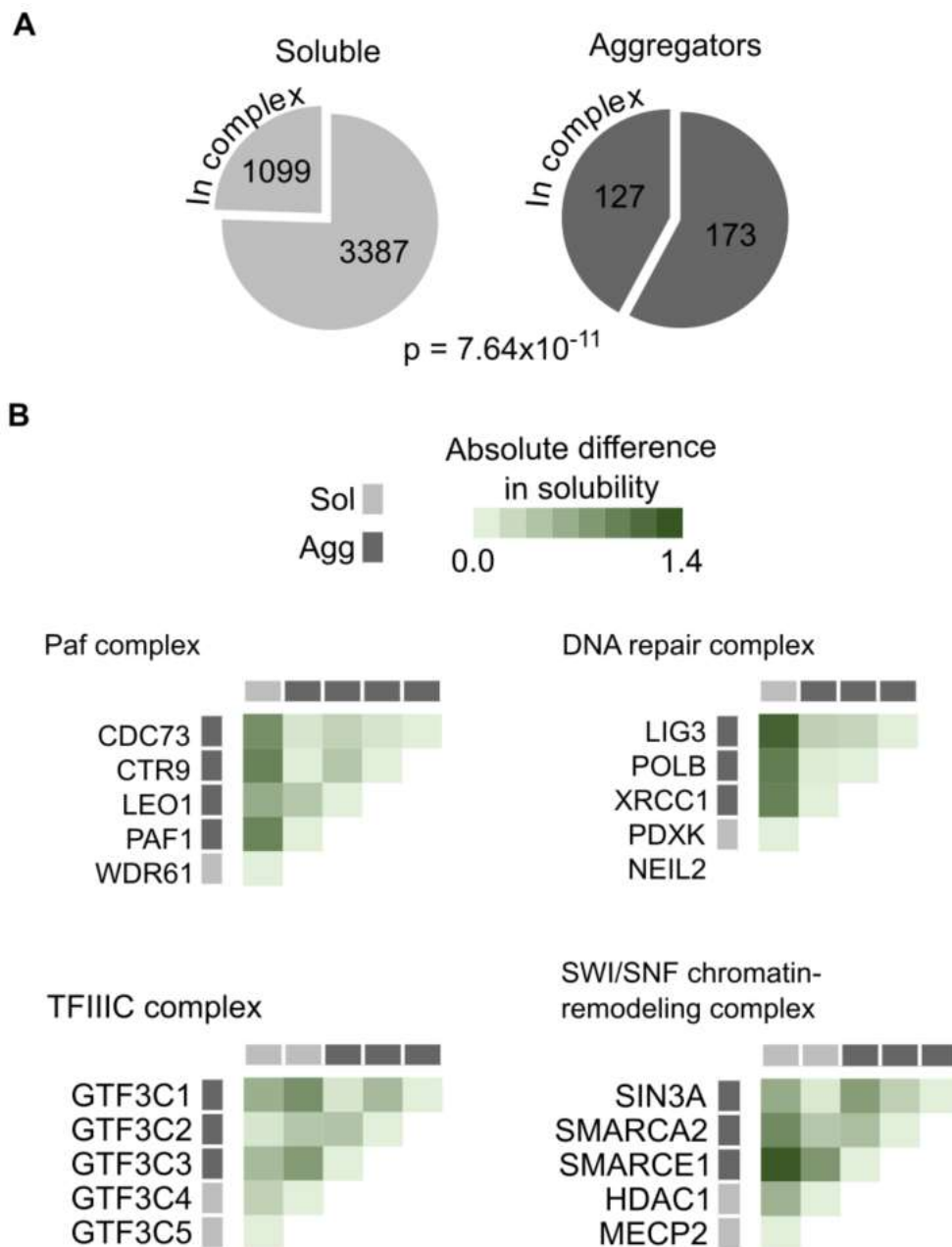
## 2.6 Protein complexes and aggregation

Being member of a protein complex could stabilize proteins (by attractive interactions between complex members) and make them less prone for aggregation. This possibility was first examined by comparing protein complex annotations between aggregating and soluble proteins and then having a more detailed analysis of individual complex members and their solubility after heat shock.

Aggregating proteins had higher proportion of proteins annotated to be a member of a protein complex than the soluble proteins (Figure 18A). From aggregators, approximately 42% were members of a protein complex (Figure 18A). On the contrary, only less than 25% of soluble proteins were complex members (Figure 18A). In other words, aggregators were enriched for protein complex members ( $p = 7.64 \times 10^{-11}$ ).

To gain a deeper view on the link between protein aggregation and protein complexes, the members of the complexes containing aggregating proteins were examined. The aggregating proteins in each complex had more similar heat shock-induced solubility changes than with soluble proteins in the same complex (Figure 18B). This indicates that the aggregating proteins might form unstable and coherently aggregating sub-structures in complexes composed of mainly aggregating proteins. To examine whether this holds true in all complexes which have at least two aggregating proteins, the average heat shock-induced solubility difference within each complex was compared to scrambled protein complexes. There was no difference observed between the average difference (Figure 19,  $p = 0.43$ ). These results indicate that the unstable and coherently aggregating sub-structure seems to be present only in complexes that have high proportion of aggregating proteins.

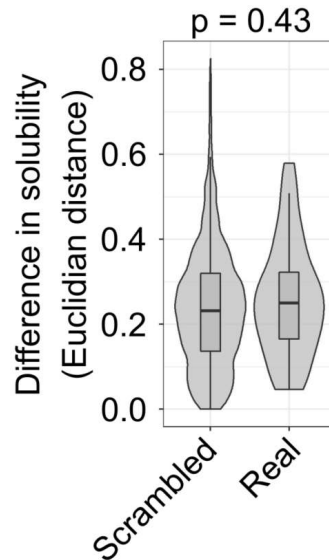




**Figure 18. Aggregators within protein complexes.**

A: Protein complex members in aggregators and soluble proteins. Pie charts showing the fraction of proteins annotated as part of a protein complex. Number of proteins are indicated inside the chart. P-value is shown for Fisher's exact test.

B: Unstable sub-structures in protein complexes that have high proportion (at least 60%) of aggregators. Heat maps showing the absolute difference in solubility between proteins within a protein complex.



**Figure 19. Solubility of aggregators within protein complexes.**

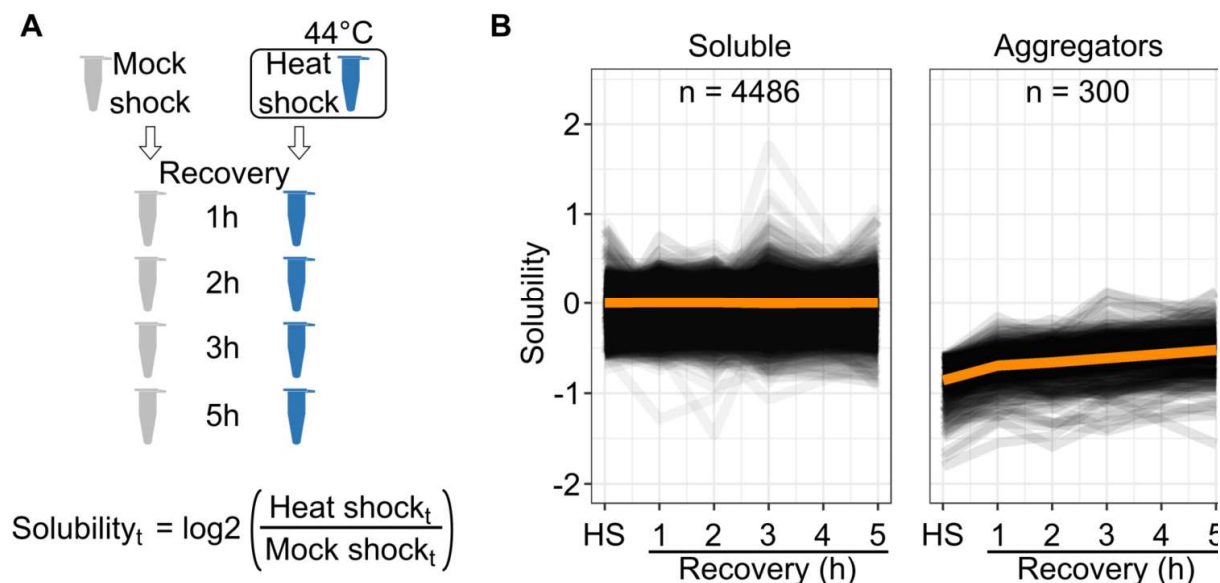
Solubility difference between aggregators within a protein complex (n = 32) and within scrambled protein complexes (n = 10,000). P-values is for non-parametric Wilcoxon test.

### 3. Protein disaggregation during recovery from heat shock

To measure protein solubility during recovery after heat shock, protein intensities were quantified from soluble fraction at each time point. For each time point, the protein intensity in the soluble fraction of heat shocked sample was compared to time-matched mock shocked control (Figure 20A). To investigate disaggregation during recovery, protein intensity only from the pre-existing fraction was measured. This allows to track the solubility of proteins that were already mature during the heat shock and exclude the effect of protein synthesis which could be wrongly interpreted as disaggregation.

#### 3.1 Wide-spread disaggregation during recovery

During the recovery, proteins from the pre-existing (light) fraction that remained soluble after the heat shock had unchanged solubility values throughout the recovery (Figure 20B). On the other hand, proteins that aggregated in the heat shock regained the



**Figure 20. Disaggregation of aggregated proteins during recovery from heat shock.**

A: Solubility measurements during recovery from heat shock. Solubility is measured as log2-transformed ratios between heat shocked and mock shocked samples at each time point.

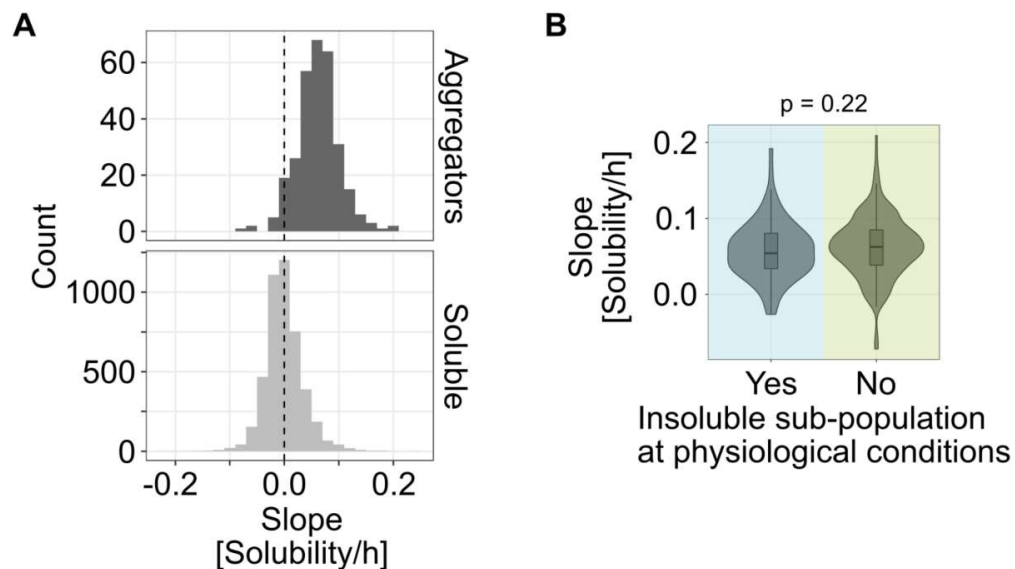
B: Solubility of proteins after heat shock. Line graphs showing protein solubility at different time points during recovery from heat shock. Data for proteins that stay soluble after heat shock (left panel) and aggregators (right panel). Each protein is shown as a line. The orange line presents the profile of an average protein in each panel.

All data shown for pre-existing protein fraction (light).

solubility during recovery (Figure 20B). Since the only source for increased protein intensity in the soluble fraction is from the insoluble fraction, this reflects wide-spread protein disaggregation. The disaggregation was not complete in the time scale of five hours. By extrapolating the linear disaggregation trend it was estimated that after approximately 12 hours a full disaggregation could be achieved.

To have a quantitative measure for the disaggregation, a linear model was fitted for each protein in Figure 20B and the slope was used as an estimate for the rate of disaggregation. This approach was valid since log2-transformed heat shock to mock shock ratios as a function of the recovery time show a linear behavior (Figure 20B).

The slopes for soluble proteins were sharply centered around zero (Figure 21A) showing an unchanged solubility as expected. The slopes for aggregating proteins had a positive shift indicative of the disaggregation (Figure 21A). The median disaggregation rate was an increase in solubility of 4.4% per hour (corresponding to 0.062 solubility units in log2-scale per hour).



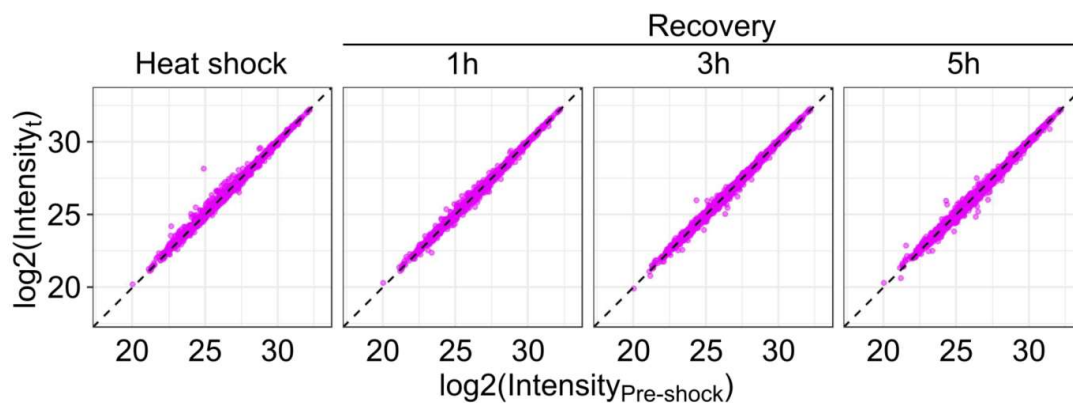
**Figure 21. Quantification of disaggregation.**

A: Quantification of solubility profiles (in Figure 20B) during recovery from heat shock. Histograms (binwidth = 0.02) showing a slope calculated for each protein.

B: Disaggregation slopes of aggregators with or without an insoluble sub-population at non-stressed conditions.

Since the presence of insoluble sub-pool in physiological conditions did not have an impact on the aggregation, it was analyzed whether it could have a role in disaggregation. If a protein is prone for forming aggregates (even in physiological condition), its further aggregation upon heat shock might make its re-solubilization slower than for the other aggregators. It was found that the disaggregation rates were very similar between proteins that did or did not contain an insoluble sub-population (Figure 21B).

Since dSILAC labeling was used (Figure 10A), the positive slope for aggregators was not because of protein synthesis. This is also evident from constant total protein amount in pre-existing fraction during the recovery time-course (Figure 22). The unchanged total protein amount also shows that there was no degradation induced for aggregators. Therefore, the increased intensity in the soluble pre-existing fraction and the constant amount of total protein present in the cells means that aggregating proteins had an increased solubility during recovery from heat shock i.e. they were disaggregated.



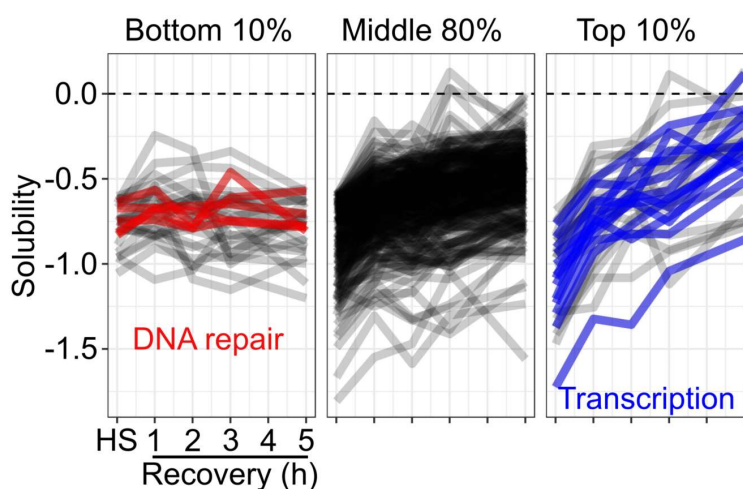
**Figure 22. Total protein levels from pre-existing fraction stays constant during recovery from heat shock.**

Total protein intensities from samples lysed with strong ionic detergent (SDS). Scatterplots comparing intensities from heat shocked samples at different time points of recovery to a pre-shock control.

### 3.2 Different disaggregation kinetics relate to different nuclear functions

The disaggregation depicted in Figure 20B appears to be quite similar for each protein. In addition, the width and shape of the slope distribution was similar between aggregators and soluble proteins (Figure 20A). This apparent homogeneity in disaggregation was analyzed in more detail. To do this, the slope values were divided to deciles. Then the proteins with the most extreme slope values were investigated in more detail.

Proteins with the lowest ten percent of the slope values were mainly not disaggregated (Figure 23). This is evident from the stable stability values throughout the recovery time of five hours. Although, it is possible that the proteins would be disaggregated later. These included many proteins related to DNA damage: tyrosyl-DNA phosphodiesterase 1 (TDP1), fanconi anemia group I protein (FANCI), DNA polymerase epsilon catalytic subunit A (POLE), telomere-associated protein RIF1 (RIF1) and protein timeless homolog (TIMELESS).



**Figure 23. Disaggregation kinetics.**

Disaggregation profiles of aggregators with the bottom 10%, middle 80% and top 10% of disaggregation slope values. Proteins related to DNA repair are highlighted (red) in the bottom 10% panel. Proteins related to transcription are highlighted (blue) in the top 10% panel.

On the other end, proteins having the highest ten percent of the slope correspond to fastest disaggregation (Figure 23). This group also consist proteins that are fully disaggregated during the time course of five hours. Approximately half of the proteins in the group were either transcription factors or proteins related to transcription (Figure 23). The transcription factors include forkhead box protein K2 (FO XK2), MAX gene-associated protein (MGA) and AT-rich interactive domain-containing protein 3A (ARID3A). The ten proteins closely related to transcription included, for instance,

transcription initiation factor TFIID subunit 4 (TAF4), transcription elongation factor B (TCEB3) and RNA polymerase II elongation factor (ELL). Therefore, the fast disaggregation was observed for proteins related to transcriptional regulation.

### **3.3 Disaggregation is driven by aggregation propensity and amount of intrinsically disordered regions**

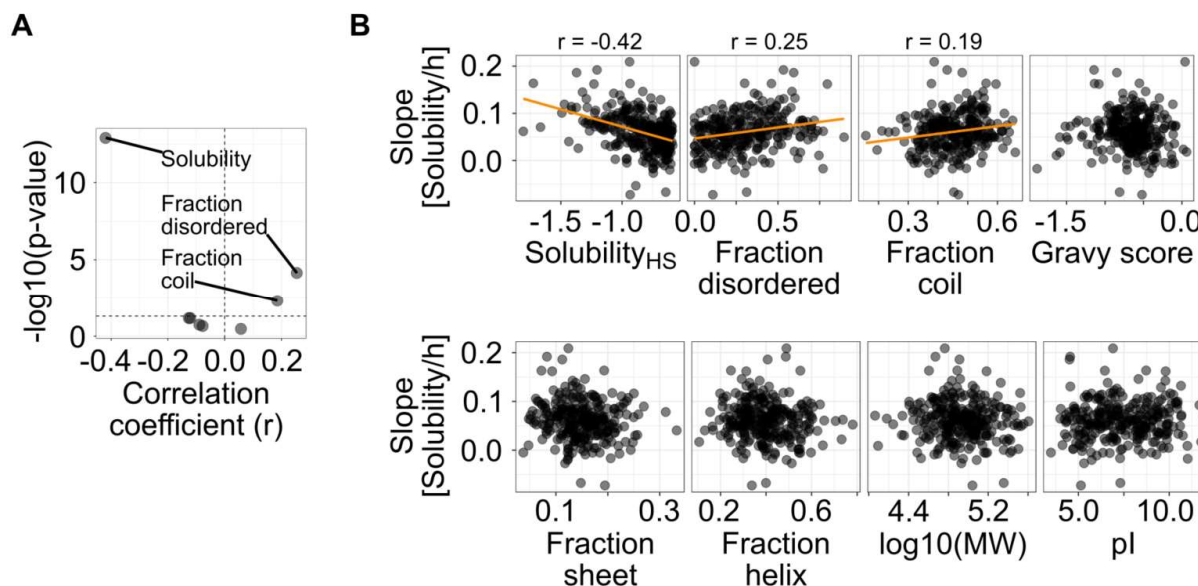
As discussed earlier, the proteins that aggregated upon heat shock were enriched for certain molecular features (Figure 16-18). Since these features were associated with aggregation propensity, they could also affect the rate of disaggregation. The disaggregation slopes of aggregating proteins were compared to the features analyzed with aggregators previously by using a correlation analysis.

A significant negative correlation was found between disaggregation slope and the loss of solubility after heat shock (Figure 24,  $p = 1.23 \times 10^{-13}$ ,  $r = -0.42$ ). This suggests that the bigger fractions of the protein pool aggregated the faster it was disaggregated. It should be noted that, as discussed earlier, that the presence of insoluble sub-population at physiological conditions did not impact neither the loss of solubility (Figure 15B) upon heat shock nor the disaggregation slope (Figure 21B). Therefore, the observed negative correlation possibly relates only to the aggregates formed upon heat shock and not the ones already present prior to heat shock.

A correlation between disaggregation slope and proportion of disordered regions ( $p = 7.13 \times 10^{-5}$ ,  $r = 0.25$ ) and random coil secondary structure ( $p = 4.97 \times 10^{-3}$ ,  $r = 0.19$ ) in a protein was found (Figure 24). Thus, the disaggregation was faster the more a protein contained disordered regions. It should be mentioned, that the proportion of disordered regions and random coil secondary structure measure similar protein properties.

Other analyzed features did not correlate with the disaggregation slope (Figure 24). Some of these features were enriched in aggregates (gravy score, fraction of beta

sheet, molecular weight and isoelectric point) while the fraction of alpha helices was not. These results suggest that other molecular features have a role in aggregation while others are related to both, aggregation and disaggregation.



**Figure 24. Molecular features of aggregators and disaggregation kinetics.**

A: Correlation analysis of disaggregation slopes and features enriched in aggregators (from Figure 16 and 18). Volcano plot showing the correlations coefficient ( $r$ ) and Benjamini-Hochberg adjusted  $p$ -values from the correlation analysis. Dashed horizontal line shows  $p$ -value of 0.05.

B: Scatterplots comparing disaggregation slopes and features enriched in aggregators. The orange trend lines and correlation coefficients ( $r$ ) are shown for features with significant (Benjamini-Hochberg adjusted  $p$ -value lower than 0.05) positive or negative correlation with the disaggregation slope.

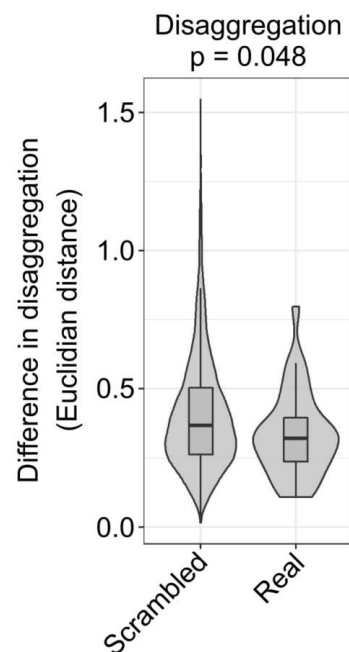
### 3.4 Protein complexes and disaggregation

Since aggregating proteins were more likely to be part of a protein complex (Figure 18A), the disaggregation was analyzed in the context of protein complexes. The disaggregation could follow similar patterns for members of the same protein complex if



the complex would be re-assembled after re-solubilization of its aggregating members. To analyze if the disaggregation of complex members was coupled, a reference dataset of scrambled protein complexes ( $n = 10,000$ ) containing randomly assigned aggregators was created. The number of aggregating proteins in each complex in the reference dataset was kept the same as in the real dataset. This means that the percentage of protein complexes that contain, for example, two aggregators is the same in both datasets.

The analysis suggests that disaggregation patterns were more similar in the real dataset than in the scrambled dataset (Figure 25,  $p = 0.048$ ). This indicates that the disaggregation was coupled to some extent between aggregators within protein complexes.



**Figure 25. Disaggregation of protein complex members.**

Differences in disaggregation slopes between aggregators within a protein complex ( $n = 32$ ) and within scrambled protein complexes ( $n = 10,000$ ). P-value is for non-parametric Wilcoxon test.

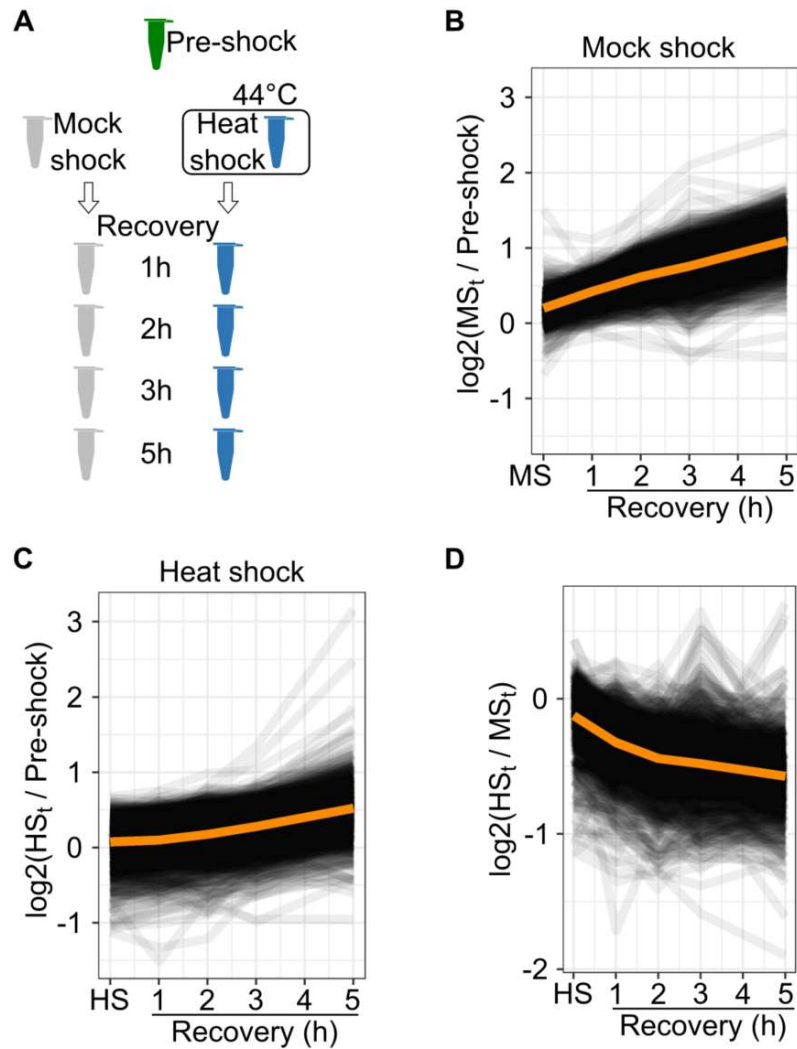
## **4. Heat shock-induced effects on protein synthesis**

The effects of heat shock to protein synthesis was examined by quantifying newly synthesized proteins (heavy) from the soluble fraction at different time points. To follow the accumulation of the heavy-labelled proteins, the protein intensity in each sample was compared to a pre-shock reference sample collected just before heat or mock shock (Figure 26A). It should be noted that since the medium switch in dSILAC was conducted 90 minutes before the heat treatment, some mature proteins already contain some heavy labelled amino acids. Therefore, some signal is already present at the first time point after heat shock. The advantage is that the heavy-labelled mature proteins increase the signal intensity to some extent and, therefore, the signal to noise ratio. This allowed to detect more proteins than if the medium switch was conducted after or right before the heat treatment.

### **4.1 Reversible stall in protein synthesis after heat shock**

After mock shock, a steady accumulation of protein intensity in the newly synthesized protein fraction was observed (Figure 26B). This corresponds to a background protein synthesis related to normal protein turnover. Since the newly synthesized protein fraction has a low intensity at the beginning (SILAC medium switch was conducted 90 minutes before mock shock), the relative increase in the signal is large, although the absolute quantity of proteins with heavy label is relatively small. This probably causes the apparent doubling of newly synthesized proteins already after five hours (Figure 26B).

After heat shock, the accumulation of newly synthesized proteins halts (Figure 26C). This is evident when comparing to the mock shock synthesis rate. However, after couple hours of recovery at 37°C, the synthesis rates recover and start to approach the levels observed in mock shocked sample (Figure 26C-D).



**Figure 26. Protein synthesis after heat shock.**

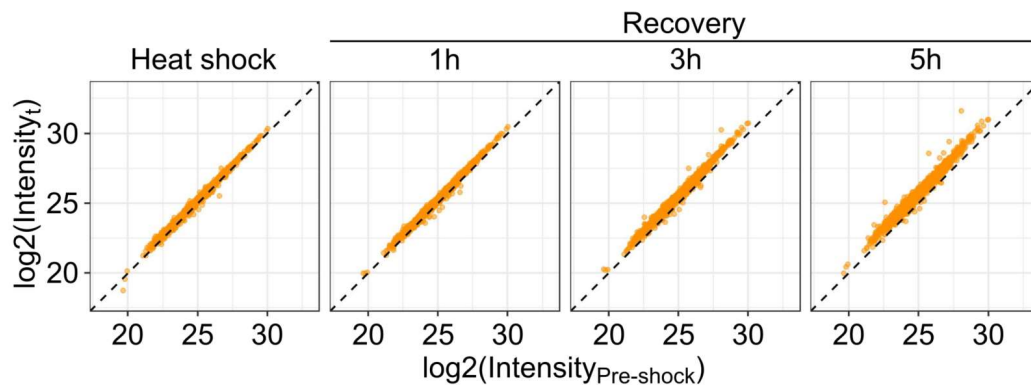
A: Schematic presentation of the samples from which newly synthesized proteins were analyzed.

B: Protein synthesis without heat shock. Line graph showing the accumulation of newly synthesized proteins after mock shock. Each protein is presented as a line. Orange line shows the profile of an average protein.

C: Protein synthesis after heat shock. Line graph showing the accumulation of newly synthesized proteins after heat shock. Each protein is presented as a line. Orange line shows the profile of an average protein.

D: Comparison of protein synthesis between heat and mock shocked samples. The ratio of protein intensities of newly synthesized proteins between heat shocked (panel C) and mock shocked (panel B) sample is shown as a line graph. Orange line shows the profile of an average protein.

The global halt in protein synthesis measured from the soluble protein fraction could also be an artefact of comprehensive aggregation of newly synthesized proteins. Therefore, the protein intensity in the total protein fraction was analyzed. The heat shock-induced stall in protein synthesis was also evident from the total protein amounts (Figure 27). The change in the levels of newly synthesized proteins was smaller when comparing them right after heat shock to one hour after heat shock than comparing the later time points. Although some aggregation can be observed (discussed next), the global halt in the accumulation of newly synthesized proteins probably stems from slowed (or completely stopped) translation rates. It should be noted that the first time point after heat shock was one hour and the halt in protein synthesis can last any time smaller than that and is beyond the scope of the experiment described here.

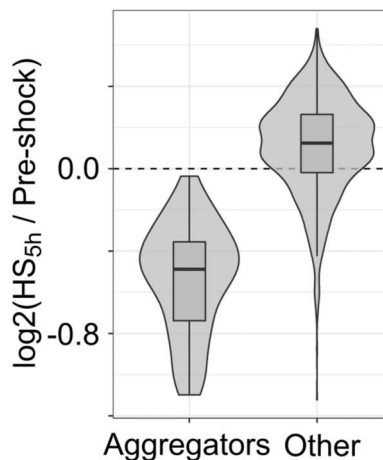


**Figure 27. Total protein levels of newly synthesized proteins during recovery from heat shock.**

Total protein intensities from samples lysed with strong ionic detergent (SDS). Scatterplot comparing intensities of newly synthesized proteins from heat shocked samples at different time points of recovery to a pre-shock control sample.

Since low amounts of proteins labelled as newly synthesized were present at the time of the heat shock, some of them aggregated (Figure 26C). These aggregating proteins could be the same as in the pre-existing fraction (mature proteins) and present aggregation-prone proteins. In addition, they could represent proteins that are unstable only at the beginning of their life cycle. To analyze this, the heat shock to pre-shock

ratio of newly synthesized proteins (from the soluble fraction) was analyzed separately for aggregators and soluble proteins (determined from the pre-existing mature proteins - i.e. light SILAC). The aggregating proteins in the newly synthesized fraction were largely the same as the ones aggregating in the pre-existing fraction (Figure 28).



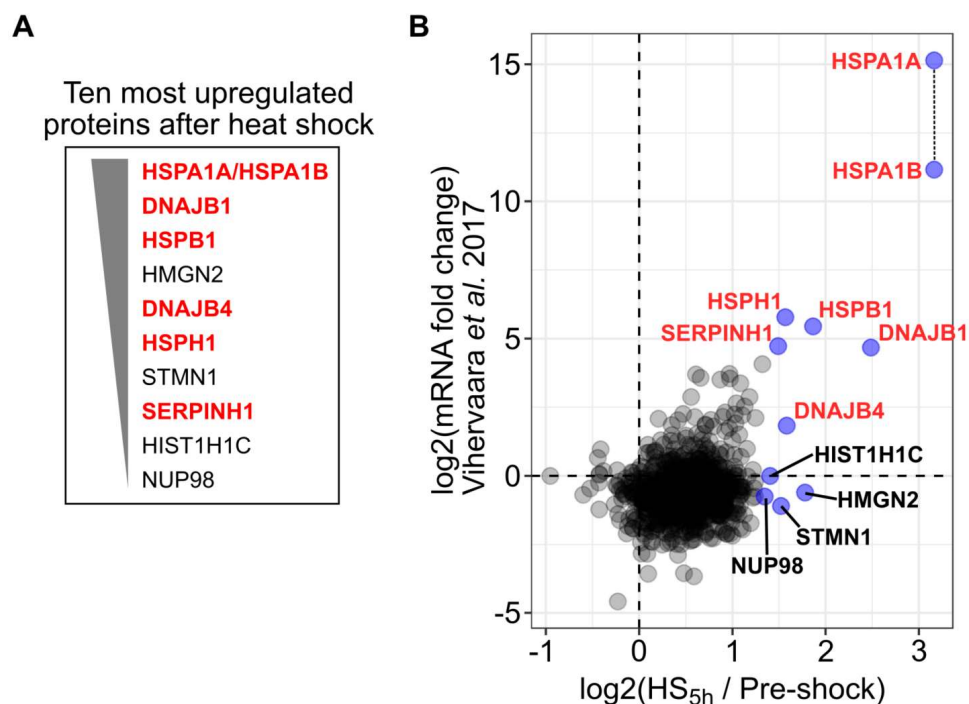
**Figure 28. Newly synthesized proteins aggregating upon heat shock.**

Combined violin- and boxplots showing the heat shock to pre-shock ratios of newly synthesized proteins. Proteins were grouped according to whether they were aggregators in the pre-existing protein fraction.

## 4.2 Upregulation of protein synthesis after heat shock

After heat shock, a strong upregulation for handful of proteins was observed (Figure 26C). This can be seen as sharp increase in the protein intensity after heat shock when compared to all other proteins - evident especially at the late time points of recovery in Figure 26C. The extent of upregulation was analyzed from the heat shock to pre-shock ratio at five hours of recovery. Taking into consideration that some newly synthesized proteins aggregated, this measure is not applicable to them; in these cases, the disaggregation and new synthesis is not distinguishable.

By examining the ten most upregulated proteins it was revealed that they included many heat shock proteins (Figure 29A). These included proteins from various major heat shock protein groups, such as Hsp70 (HSPA1A-HSPA1B), Hsp40 (DNAJB1 and DNAJB4) and small heat shock proteins (HSPB1). Other heat shock proteins were HSPH1 (Hsp105) and SERPINH1 (Hsp47). The high sequence similarity between the paralogs HSPA1A and HSPA1B caused them to be indistinguishable in the MS analysis; the results probably reflect a combination of the two. The other proteins in this



**Figure 29. Upregulation of protein synthesis after heat shock.**

A: Proteins with the strongest upregulation after heat shock. Translation rates were estimated from the ratio between protein intensity at five hours after recovery to the pre-shock control (last time point in Figure 26C). Ten most upregulated proteins are listed in a decreasing order. Heat shock proteins are highlighted in bold red.

B: Comparison of protein level regulation and transcription level regulation after heat shock. Scatterplot comparing protein synthesis five hours after heat shock to mRNA levels after heat shock [adapted from (Vihervaara *et al.*, 2017)]. The top ten upregulated proteins (from panel A) are annotated and shown in blue. Heat shock proteins are highlighted with bold red. HSPA1A and HSPA1B could not be distinguished in the mass spectrometry analysis; they are plotted as they would have the same intensity in the protein data (depicted with vertical dashed line).

group include non-histone chromosomal protein (HMGN2), stathmin (STMN1), histone H1.2 (HIST1H1C) and nuclear pore complex protein NUP9.

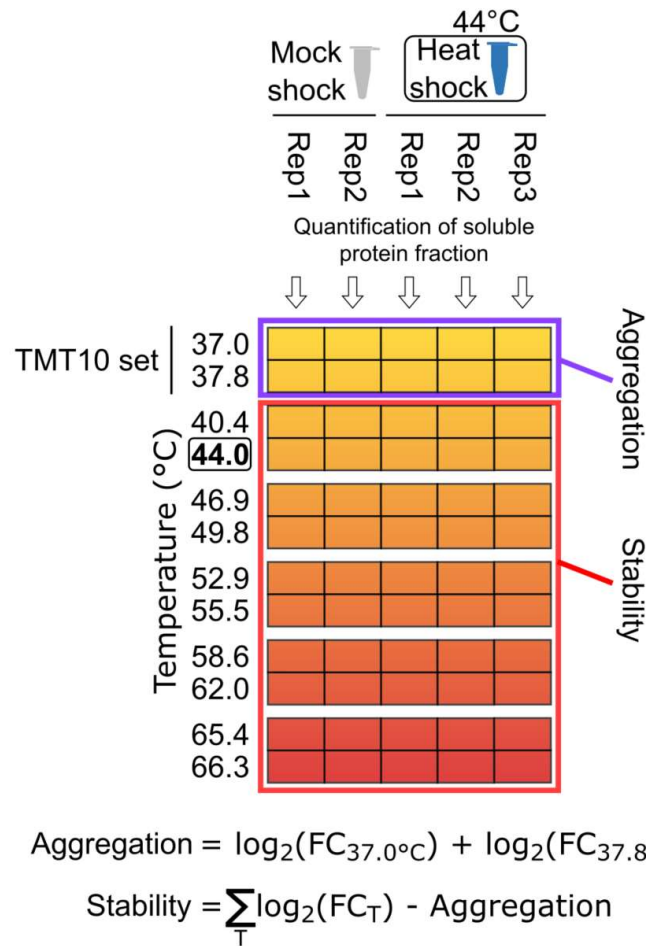
The regulation of protein synthesis can be done on many levels ranging from transcription, mRNA processing to translational control. To analyze a possible connection between the changed levels of mRNA and proteins upon heat shock, the protein level upregulation was compared to a previously published study of transcriptional regulation after heat shock (Vihervaara *et al*, 2017). In the transcriptomics study, the same cell line as here (K562 cells) was used with a heat shock of 42°C for 30 minutes (here, 44°C and ten minutes) and changes in mRNA levels were quantitatively analyzed. The reported change in mRNA levels and protein levels (measured after five hours of recovery) were then compared.

It was found that by large the changes in mRNA and protein level did not correlate (Figure 29B). However, the most upregulated mRNAs did have a correlation with the protein level upregulation. Furthermore, from the proteins with the highest protein level upregulation, the mRNA levels had a strong upregulation only with the heat shock proteins while the non-heat shock proteins were upregulated only at the protein level (Figure 29B). This suggests that for some proteins, the regulation of expression is translation rather than transcription driven. Interestingly, similar results were observed with yeast (Muhlhofer *et al*, 2019).

## **5. Thermal stability changes of soluble proteins**

Proteins aggregate upon heat shock (and other stress conditions). The aggregation and disaggregation was discussed in the previous sections. To extend the analysis, heat shock-induced effects on proteins that stayed soluble were analyzed. To do this, the thermal stability of those proteins were analyzed with two-dimensional thermal proteome profiling (2D-TPP) (Becher *et al*, 2016, Mateus *et al*, 2017, Savitski *et al*, 2014). In essence, the aggregation propensity of proteins is compared between heat

shocked and mock shocked samples in a proteome-wide thermal shift assay. In the assay, samples are distributed to aliquots that are exposed to temperature gradient that goes well above (up to 66.3°C) the 44°C used in the heat shock experiment. These temperatures eventually denature and aggregate almost all proteins in the cell.



**Figure 30. Schematic presentation for two-dimensional thermal proteome profiling.**

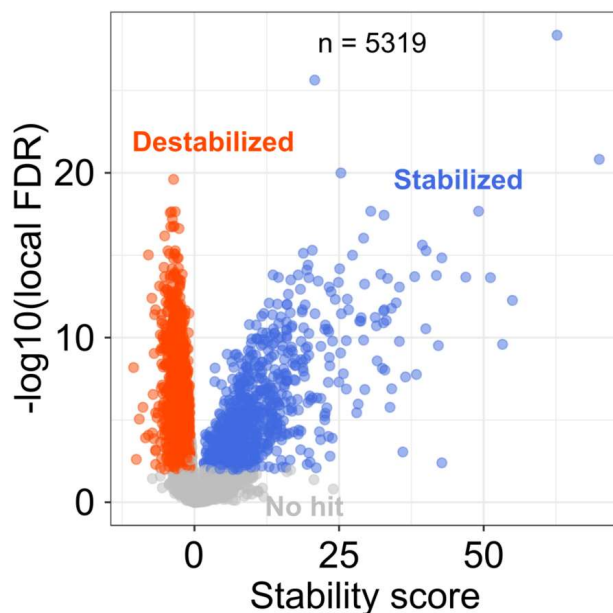
Replicates of heat shocked and mocked shock samples were aliquot and exposed to a gradient of temperatures that reach beyond the 44C used in the heat shock, eventually denaturing all proteins in the samples. Soluble protein fraction of each aliquot is collected, digested to peptides and labelled with TMT tags. The labelling was conducted to pool heat shocked and mocked shocked samples from two adjacent temperatures ('TMT 10 set' exemplified for the two lowest temperatures). The ratio of protein intensity in the soluble fraction was compared between heat shocked and mock shocked samples and adjusted for the possible aggregation upon heat shock. These ratios are converted to stability scores (see materials and methods) that capture the change in thermal stability induced by the heat shock.



The intensities of soluble proteins were measured from heat and mock shocked samples and their levels were compared within each temperature in the 2D-TPP assay (Figure 30). After correcting for possible aggregation in the heat shocked samples, the heat shock to mock shock ratios at each temperature are summarized to a stability score (Figure 30). High stability score refers to higher thermal stability in the heat shocked samples and *vice versa*.

### 5.1 Thermal stabilization of soluble remnants of aggregators

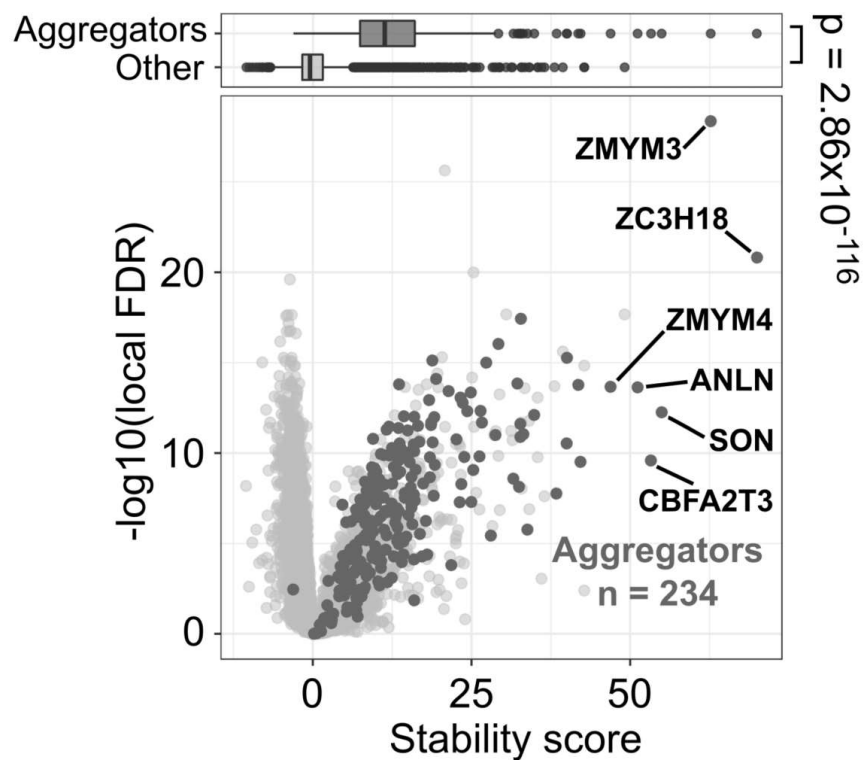
A dataset containing thermal stability data for 5319 proteins quantified from at least six different temperatures with minimum of two unique peptides was acquired. Heat shock induced wide-spread changes in thermal stability of proteins, both destabilization and stabilization (Figure 31). The measured effect of stabilization was stronger than for destabilization as indicated by higher absolute changes in the stability score.



**Figure 31. Heat shock-induced changes in thermal stability of soluble proteins.**

Scatter plot showing the stability scores. Destabilized proteins are highlighted in red and stabilized proteins are highlighted in blue.

When analyzing the stabilized protein, it was found that they contained many proteins that aggregated in the heat shock (Figure 32). As mentioned earlier, thermal stability can be measured for proteins that are soluble. This apparent paradox is explained by the fact that protein aggregation is not complete. As discussed previously, the highest loss of solubility within aggregators was 75%. A soluble sub-pool remains while a significant fraction forms aggregates. The results show that this soluble sub-pool is more thermally stable since it required higher temperatures to aggregate. Interestingly, small ubiquitin-related modifier 1 (SUMO1) was the only aggregator that was destabilized. It would be tempting to suggest that the one destabilizing protein would be a false positive.



**Figure 32. Soluble remnants of aggregators stabilized upon heat shock.**

Scatter plot for stability scores. Soluble remnants of aggregators are highlighted (dark grey) and the most strongly stabilized are annotated. Box plots (top) shows stability scores for aggregators and all other proteins separately. P-value (top) is shown for non-parametric Wilcoxon-test.

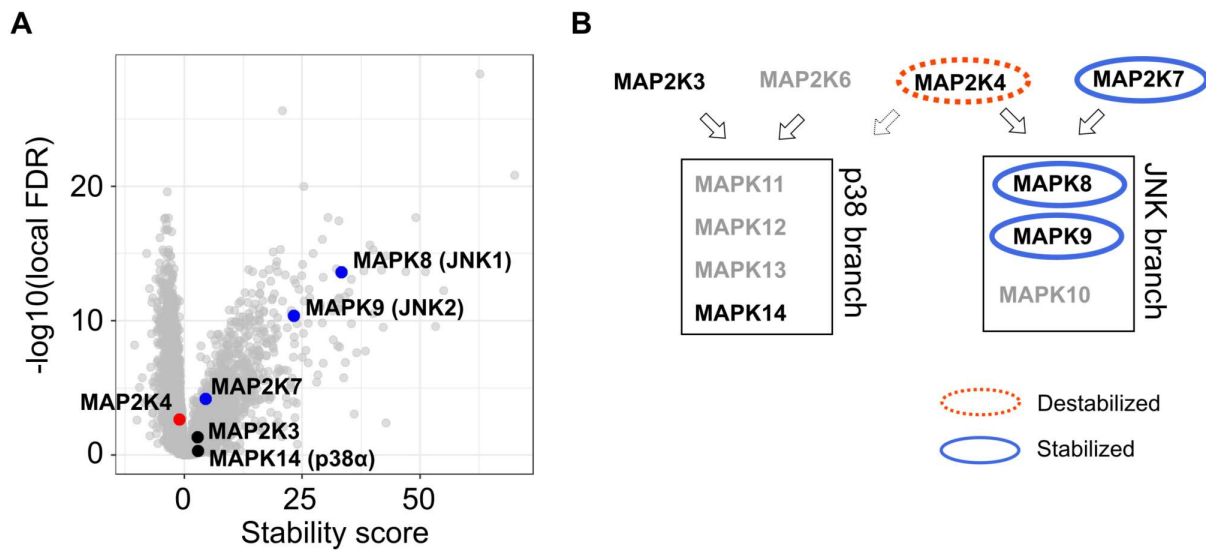
Among the soluble remnants of aggregators, the strongest thermal stabilization was measured for zinc finger CCCH domain-containing protein 18 (ZC3H18) and zinc finger MYM-type protein 3 (ZMYM3) (Figure 32). Interestingly, as the names suggest, both of them contain a zinc-finger domain. In addition, zinc finger MYM-type protein 4 (ZMYM4), another zinc-finger containing protein was also among the most stabilized aggregators (Figure 32).

## **5.2 Thermal stability changes and stress signaling**

Among the proteins that remained soluble after heat shock, some of the strongest stabilizations were observed for MAPK8 and MAPK9 (Figure 33). The two proteins are mitogen-activated protein kinases (MAPK) which belong to a c-Jun N-terminal kinase (JNK) group and constitute a branch in stress-activated MAPK pathway (Davis, 2000, Hotamisligil & Davis, 2016) (Figure 33).

As the name suggests, stress-activated MAPK pathway is a signaling pathway activated in environmental stress or by inflammatory cytokines (Hotamisligil & Davis, 2016). It has two branches involving MAPKs from either JNK or p38 group (Davis, 2000, Hotamisligil & Davis, 2016) (Figure 33). These kinases also have specific upstream kinases that regulate their activity (Davis, 2000, Hotamisligil & Davis, 2016) (Figure 33). Heat shock-induced changes in thermal stability was observed only in JNK branch, including JNKs themselves as well as their upstream kinases (Figure 33). This specific stabilization of only one branch might reflect its activation upon heat shock. When activated, JNK kinases can have a dual role in responding to stress. On the one hand, they promote cell survival response but prolonged activation can lead to apoptosis or necrosis (Hotamisligil & Davis, 2016)

.



**Figure 33. Heat shock-induced stability changes in stress-activated MAPK pathway.**

A: Volcano plot highlighting stability changes among all members of stress-activated MAPK pathway. Destabilized proteins are highlighted in red and stabilized proteins are highlighted in blue.

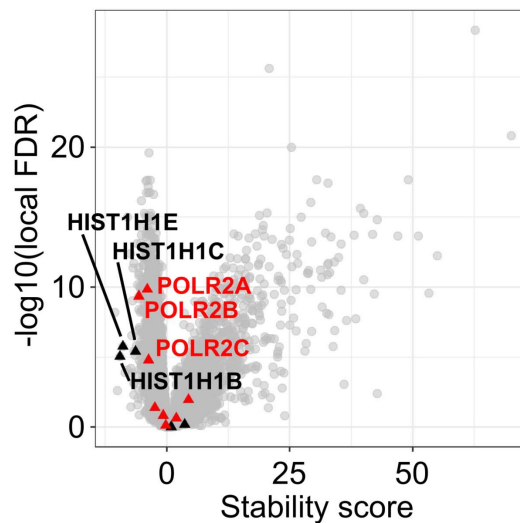
B: Representation of stress-activated MAPK pathway [adapted from (Davis, 2000) and (Hotamisligil & Davis, 2016)]. Boxes present the two branches composed of p38 and JNK kinases. Their upstream kinases are indicated at the top. The arrows show the target branch for each kinase. The activation of p38 branch by MAP2K4 is unclear (Hotamisligil & Davis, 2016) and is shown as an arrow with dashed outline.

### 5.3 RNA polymerase II destabilized upon heat shock

After heat shock, a global down-regulation of transcription has been observed (Vihervaara *et al*, 2017). In humans (and other eukaryotes), transcription is mediated by RNA-polymerase-II (Pol-II) (Kornberg, 1999, Sims *et al*, 2004). Destabilization of Pol-II subunits has been previously shown to result from DNA detachment (Becher *et al*, 2018). The thermal stability of many Pol-II subunits was significantly lowered in heat shock (Figure 34).

## 5.4 Destabilization of H1 histones

H1 histones are proteins that bind nucleosomes together in compact DNA (Hergeth & Schneider, 2015). The H1 histones were among the most destabilized proteins after heat shock (Figure 34). The C-terminus of H1-histones is unstructured when not bound to DNA (Roque *et al*, 2005). Therefore, the heat shock-induced destabilization of H1 histones might reflect detachment from DNA and therefore opening of compact DNA. In accordance with these results, detachment of Hhop1 (H1 histone homolog in yeast) from DNA upon heat shock has been described earlier (Zanton & Pugh, 2006).



**Figure 34. Destabilization of proteins related to DNA binding.**

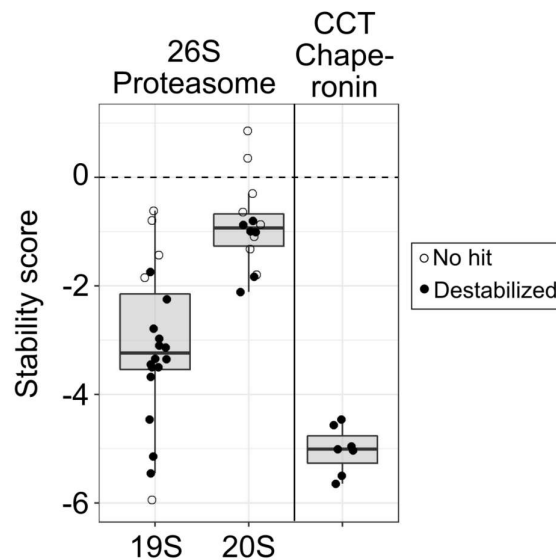
Volcano plot of stability scores highlighting members of RNA polymerase II (red triangles) and H1 histones (black triangles). Destabilized proteins are labelled.

## 5.5 Comprehensive destabilization of quality control protein complexes

Changes in thermal stability was observed for 26S proteasome and chaperonin containing TCP-1 (CCT) chaperonin complex (Figure 35). Both of them are crucial components in protein quality control. CCT chaperonin assists in *de novo* protein folding

and it is estimated that approximately 10% of proteins need CCT for folding (Lopez *et al*, 2015). 26S proteasome on the other hand degrades proteins and is therefore an essential part quality control (Coux *et al*, 1996).

26S proteasome complex is composed of 19S regulatory particle and 20S core particle (Coux *et al*, 1996). The destabilization was stronger for 19S regulatory particle than for the 20S core particle (Figure 35). In an experiment where ATP was added to cell lysates, 19S regulatory particle was thermally stabilized (Sridharan *et al*, 2019). This would suggest ATP depletion in heat shock. However, ATP levels were unchanged (Figure 9B; the viability measurements are based on ATP levels). The strong destabilization of 19S regulatory particle could reflect its impaired capability to bind ATP. It has been shown that ATP-driven proteasomes are impaired in acute heat shock (Kuckelkorn *et al*, 2000).



## DISCUSSION

In this project, a mass spectrometry-based platform was developed to study protein aggregation, disaggregation and protein synthesis in the context of a non-lethal heat shock in human cells *in situ*. The method includes a dynamic SILAC protein labelling in cell culture which enables to distinguish between pre-existing mature and newly synthesized proteins. Sample labelling with TMT at peptide level was used in combination with the dynamic SILAC to allow sampling of multiple time points during recovery from the heat shock. Sample processing with non-ionic mild detergent (NP-40) allowed to analyze relative protein solubility after heat shock and during multiple time points of recovery.

It was found that aggregating proteins contained high proportion of intrinsically disordered regions and were predicted to contain more random coil secondary structure than proteins that stayed soluble in heat shock. In addition, aggregated proteins were enriched in high molecular mass, hydrophilic amino acids and are mainly annotated to be nuclear proteins. Aggregators were also enriched for protein complex members. Therefore, endogenous human proteins have differences in their propensity for aggregation in heat stress which are based on intrinsic physicochemical features as well as cellular compartmentalization.

The majority of aggregating proteins disaggregated during recovery from heat shock. Disaggregation rates correlated with the aggregation propensity, the amount of intrinsically disordered regions and predicted random coil secondary structure. Other features enriched in aggregating proteins did not correlate with the disaggregation rates. Few proteins showed no evidence of disaggregation, including proteins related to DNA repair. The fastest disaggregating proteins contained functions related to transcriptional regulation. These results show that disaggregation has a major role in recovery from

heat stress. In addition, the disaggregation (or absence of it) is possibly tuned for substrates involved in different cellular functions.

After heat shock, the protein synthesis stalled for approximately two hours. However, a gradual increase in the synthesis rate was observed at the later time points which eventually reached the control levels at the late time points of the recovery. Few proteins showed a marked increase in the synthesis rates. These proteins included many heat shock proteins. The protein level upregulation correlated with the upregulation of mRNA levels only for the most upregulated transcripts.

In addition to the hyperplexed platform combining dSILAC and TMT labelling, changes in thermal stability of proteins that remained soluble after heat shock was analyzed with two-dimensional thermal proteome profiling. Stabilization of soluble remnants of aggregators was found. This revealed that aggregators have a sub-pool of proteins that were resistant to heat shock-induced aggregation. Stability changes were found also for protein species that stayed soluble upon heat shock, including stress-activated MAPK pathway, Pol-II subunits, H1 histones and quality control complexes. These stability changes probably have a functional link to their activities.

## **1. Nucleus is sensitive for proteotoxic stress**

The localization of aggregating proteins was analyzed with Human Protein Atlas annotations. These annotations are based on experimental microscopy that uses antibodies to detect and localize proteins. By using the data with highest supportive evidence, the aggregating proteins were more often associated with localization terms related to nucleus than soluble proteins (Figure 13). This is systematically true for all



nuclear annotations. At the same time, all major cytoplasmic compartments were more often associated with soluble proteins. This quite clear separation of nuclear and cytoplasmic proteins to aggregating and soluble proteins suggests that nuclear proteins are the most sensitive for heat stress. It should be mentioned that the localization annotations are done at physiological conditions in the absence of stress. Protein localization can change in stress. For example, many proteins translocate from nucleolus to cytoplasm (and from cytoplasm to nucleus) upon DNA damage induced by etoposide (Boisvert & Lamond, 2010). As the localization in the nucleus or cytoplasm can have drastic effect on protein (thermal) stability, it could be speculated that the aggregation of nuclear proteins could stem from stress-induced translocation and subsequent destabilization; the nuclear localization would therefore not have a contribution to the protein stability. However, there is no evidence for such a massive shuttling of proteins between nucleus and cytoplasm upon heat stress.

In addition to the analysis based on localization annotation, GO term enrichment analysis also showed similar results. The GO terms enriched in aggregating proteins contained almost exclusively terms related to nuclear functions (Figure 12). These included DNA binding, nucleolus and chromatin remodeling, for example.

Furthermore, the strong enrichment of intrinsically disordered regions in aggregators could be linked to nuclear localization since DNA-binding proteins are known to contain high proportion of disordered regions (Fuxreiter *et al*, 2011, Vuzman & Levy, 2012). In accordance, GO terms related to DNA-binding were enriched in aggregating protein (Figure 12).

These results discussed above suggest that the proteome of the nucleus in human cells is sensitive for proteotoxic stress. In this light, it is note-worthy that upon heat shock,

accumulation of Hsp70 chaperone system to nucleolus (Frottin *et al*, 2019, Welch & Feramisco, 1984) and nuclear speckles (Deane & Brown, 2017) in human cells has been reported. The need for a Hsp70 chaperone machinery in these nuclear organelles might also suggest that proteins aggregate to these sites rather than as “free” aggregates in the nucleoplasm. In fact, misfolded proteins have been reported to accumulate in to nucleolus upon heat stress (Frottin *et al*, 2019). The limitation of the experiment described here is that it provides no direct information about the protein localization. Extending the scope with imaging mass spectrometry could add more information to the protein localization upon heat shock (and during recovery) while still allowing proteomics analysis of endogenous proteins.

Since nuclear proteins contain higher amounts of intrinsically disordered regions, it has been thought that they would be less stable than cytoplasmic proteins. Although the concept of unstable nuclear proteins has emerged from these findings (Jones & Gardner, 2016), it has been accepted without comprehensive and systematic experimental evidence. Thus, the results described here offers the first proteome-wide data that supports this view.

Interestingly, mass spectrometry-based aggregation studies conducted with yeast have not identified nuclear proteins enriched in aggregating proteins (Ibstedt *et al*, 2014, O'Connell *et al*, 2014, Wallace *et al*, 2015, Weids *et al*, 2016). In the light of the results presented here, the absence of any nuclear related terms in the functional enrichments (such GO term enrichment performed here) reported in these studies is interesting. Although, nucleolus was one enriched term in the context of heat shock (Wallace *et al*, 2015) and nuclear transport in the context of arsenite stress (Ibstedt *et al*, 2014). If disordered regions would make nuclear proteins less stable, why it is not evident from multiple experiments done with yeast?

## 2. DNA-binding proteins contribute to the nuclear sensitivity

As mentioned earlier, DNA-binding proteins were enriched in aggregating proteins in the GO term enrichment analysis (Figure 12). In addition, the high proportion of disordered regions in aggregating proteins can be linked to DNA-binding proteins (Fuxreiter *et al*, 2011, Vuzman & Levy, 2012). This leads to a question that how much DNA-binding proteins contribute to the aggregation propensity of nuclear proteins.

Recent proteome-wide studies mapping melting points of human and bacterial proteome offer a way to compare the stability of proteins within each proteome. In human K562 cells, proteins related to DNA-binding proteins were found to have the lowest melting points, i.e. they are the least (thermally) stable proteins in the proteome (Savitski *et al*, 2014). In bacteria, the proteins with the lowest melting points were topoisomerases (proteins that unwind DNA) and proteins related to DNA replication (Mateus *et al*, 2018). Together these results point towards DNA-binding proteins - or proteins closely related to DNA - as the most sensitive proteins to contribute to the nuclear sensitivity.

If the DNA-binding proteins are more prone for aggregation than other proteins, their disaggregation during recovery was the fastest (Figure 23). It could be speculated that these observations are related to the important cellular function mediated by DNA-binding proteins, e.g. transcription and gene regulation. Quick sequestration to aggregates upon stress could protect them from toxic interaction. After the stress is over, these proteins are quickly needed for normal or rewired gene expression and hence are re-solubilized fast. It remains an open question how much the aggregation of DNA-binding proteins relates to the transcriptional arrest observed in heat shock (Vihervaara *et al*, 2017).

### 3. Aggregation and disaggregation of protein complexes.

It was found that aggregating proteins contained more protein complex members than the soluble proteins (Figure 18A). These complexes were related to DNA binding (Figure 18B). The stabilities of protein complex members are linked to the integrity of the complex. If the energy input from the heat shock is enough to interfere with the inter-molecular interactions between complex members, it could cause the complex members to unfold and aggregate. However, it would be specific for complexes that bind to DNA (or chromatin in general). Therefore, it is challenging to separate the two possibilities - aggregation due to DNA-binding (as discussed in the previous section) or being a complex member - making these proteins prone for aggregation.

Members of protein complexes did not aggregate coherently (Figure 19). This suggests that the complexes break in heat shock and the extent of aggregation is related to the individual properties of each protein species. On the other hand, the melting of protein complex members has been reported to be coherent (Becher *et al*, 2018, Tan *et al*, 2018). This could suggest that in (mild) heat shock protein complexes break apart and the aggregation is dependent on the characteristics of each member, while the melting (including heat treatments on a wide range of temperatures) is more consistent especially at higher temperatures.

In contrast to aggregation, a slight trend towards coherent disaggregation of protein complex members was observed (Figure 25). If aggregation of protein complex members is not coherent, the coherent disaggregation suggests that complex members are specifically chosen from the aggregates and disaggregated together. This would imply a very specific action for disaggregase since the complex members to be disaggregated together are not aggregated coherently. The possible coherent disaggregation could also be explained by selective aggregation of the complex members to the same deposit (although, not as an intact protein complex).

#### 4. Protein aggregation to quality control sites in nucleus

A nuclear aggregation site - intranuclear quality control compartment - has been described in yeast (Miller *et al*, 2015). Similar site has not been found in human cells. Therefore, it is interesting to speculate if and how protein aggregation is organized in the nucleus of human cells. The presence of unstable proteome in nucleus of human cells could suggest that the human nucleus would contain an aggregation deposit site as well.

Recent reports highlight that stress-induced protein aggregation can happen on specific nuclear sites in human cells. Accumulation of misfolded proteins has reported to take place on nucleolus upon heat shock (Frottin *et al*, 2019). In addition, a large number of proteins had an increase in abundance on chromatin after heat shock (Aprile-Garcia *et al*, 2019) - although, the authors explored it in the context of transcriptional regulation. Moreover, nuclear stress bodies that include proteins related to transcription and splicing form in stress (Biamonti & Vourc'h, 2010). Although, together these examples do not suggest the presence of a unique deposit site, but they highlight that aggregation in nucleus can be organized.

Interestingly, nuclear stress bodies might be quite specific for higher organisms, since they were not identified from rodents (Jolly & Lakhota, 2006). This might explain why aggregates formed during different stresses in yeast do not contain an over-representation of nuclear proteins. The GO terms enriched in aggregating proteins (Figure 12) also overlap with the proteins reported to form nuclear stress bodies (Biamonti & Vourc'h, 2010); these include mainly proteins related to transcription and chromatin remodeling. In this view, the aggregation could relate to protein sequestration in order to perform a certain cellular function rather than depositing damaged proteins to a quality control site.

The sequestration of misfolded proteins to nucleolus was linked to its function as a protein quality control compartment in nucleus, since disruption of the formation of nucleolus led to delayed processing of misfolded proteins (Frottin *et al*, 2019). Interestingly, the proteins accumulating in nucleolus upon heat shock were enriched in disordered regions (Frottin *et al*, 2019). This suggests that the aggregation nucleus could mean sequestration to nucleolus. However, the endogenous proteins with the reported increase in abundance at nucleolus upon heat shock (Frottin *et al*, 2019) contained mainly nucleolar proteins and the reported cytosolic proteins were mainly ribosomal proteins; nucleolus as an assembly site naturally contain ribosomal proteins (annotated cytosolic). This suggests that the protein accumulation in nucleolus upon heat shock would actually reflect re-arrangement or morphological changes of nucleolus. Therefore, it remains an open question how much nucleolus contributes to the sequestration of disordered endogenous proteins.

The fraction of aggregating proteins that remained soluble in heat shock were thermally stabilized (Figure 32). It could be speculated that it is related to heat shock-induced binding to DNA (or “aggregation” on chromatin) (Aprile-Garcia *et al*, 2019). Although, further experiment would be needed to find causes behind thermal stabilization, DNA binding has been shown to stabilize subunits of RNA polymerase II (Becher *et al*, 2018). Interestingly, the most stabilized proteins contained zinc-finger domain containing DNA-binding proteins (Figure 32). Since aggregating proteins were enriched in DNA-binding proteins and are stabilized upon heat shock, it would suggest that DNA-binding might have a role in it. It is worth mentioning that the stabilized proteins might reflect a sub-population that was already more stable before the heat shock or the stabilization was induced by the heat shock. In the former case, these proteins could be bound to DNA which would cause them to be more stable. In the latter case, the DNA binding would be a response to the heat shock which would open up interesting scenarios. As mentioned before, the accumulation of proteins to chromatin upon heat shock has been demonstrated (Aprile-Garcia *et al*, 2019).

Assuming that DNA-binding proteins would respond to heat shock by binding to DNA, it opens a question if chromatin binding could be a mechanism to protect, for example, a small but essential pool of sensitive DNA-binding proteins from uncontrolled aggregation. This pool would then be readily available once the stress is over. It could also reflect the formation of nuclear stress bodies. In this case, the response might not relate to protecting the unstable proteins but rather suggests that these proteins would have important functions in the stress response. It could be speculated further that the lower stability of these proteins would make them more easy to sequester into deposit sites such as nuclear stress bodies.

## **5. Intrinsically disordered regions protect from toxic interactions and facilitate disaggregation**

Intrinsically disordered regions were found not only to be highly enriched in aggregating proteins (Figure 17A) but they also facilitated disaggregation (Figure 24). A similar result was found for predicted random coil secondary structure (Figure 17C and 24B). These two measures are related since they analyze a similar property, although, from somewhat different angles. It has been reported that proteins containing intrinsically disordered regions are more prone for aggregation *in vitro* (Uemura *et al*, 2018) and *in vivo* (Hosp *et al*, 2017, Walther *et al*, 2015). It should be noted, that the aggregation of proteins with disordered regions can mean loss of solubility (Hosp *et al*, 2017, Walther *et al*, 2015) as well as formation of phase separated membrane-less organelles (Bolognesi *et al*, 2016). It has been suggested that these proteins are actively sequestered to avoid uncontrolled aggregation, since proteins with intrinsically disordered regions are associated with many protein aggregation diseases (Walther *et al*, 2015). Therefore, the potential toxicity of proteins with disordered regions would favor their isolation from other proteins by sequestering them in aggregates. During heat shock, the whole proteome is exposed to the stress; in addition, chaperones and other

members of the quality control might be overwhelmed and unable to protect all proteins prone for misfolding (and aggregation). The sequestration could protect disordered proteins from acquiring toxic conformations induced by interactions with other misfolded proteins; on the other hand, sequestering the disease-prone proteins could protect other misfolded proteins from the potential toxicity of disordered proteins.

It would be tempting to conclude that intrinsically disordered regions have properties that can make proteins prone for aggregation. This could stem from the physicochemical properties of disordered regions or active sequestration of them as potentially toxic proteins. However, removing clearly defined disordered regions did not make proteins less sensitive for aggregation *in vitro* (Uemura *et al*, 2018). This would suggest that the physicochemical properties of disordered regions do not contribute to their instability. In addition, disordered regions themselves are soluble (Uemura *et al*, 2018). This suggests that the aggregation driven by the hydrophobic effect does not seem plausible. Therefore, it could be speculated that disordered regions could be an adaptation to aggregation-prone proteins rather than the cause behind aggregation.

Having disordered regions in aggregation-prone proteins could have some benefits. In the light of evolution, their presence probably reflects some fitness gain related to them - especially when considering their potential toxicity. They could, for example, shield toxic interactions between aggregating proteins (such as formation of amyloid fibers). Owing to their conformational flexibility, the disordered regions could buffer the spread or formation of low energy conformations. This could be achieved by interfering and disabling the formation of  $\beta$ -sheets prevalent in amyloid fibers. Interestingly, a cytoprotective role has been described for small heat shock proteins that contain disordered N- and C-terminal domains flanking an alpha crystalline-domain (Mogk *et al*, 2019). They are important in sequestering proteins to aggregates - and aggregation sites (Specht *et al*, 2011). In addition, disaggregation of the aggregate “coated” with small heat shock proteins is more efficient (Zwirowski *et al*, 2017). Proteins aggregating with small heat shock proteins might be better able to retain their native (or close to



native) fold (Mogk *et al*, 2018). This would protect the proteins from gaining toxic conformations and fasten their refolding process after disaggregation.

If intrinsically disordered regions could cushion the interactions between aggregating proteins, it could also explain the observed faster disaggregation rates associated with high amount of disordered regions and predicted random coil secondary structure (Figure 24). The interactions between aggregating proteins would not be so strong and the energy needed for disaggregation would be lower. In addition, high amounts of flexible disordered parts could provide flexible loop regions in the aggregates for disaggregases to act up upon. This would add more surface area available for disaggregase enabling the recruitment of higher number of them.

## **6. The aggregation of proteins with high molecular weight, hydrophilic character and high pI**

The finding that proteins with high molecular weight are prone for aggregation (Figure 17B) has been also reported with experiments done on different organisms beside human. For example, in yeast (Weids *et al*, 2016) and mice (Hosp *et al*, 2017), aggregates had enrichment in high molecular weight proteins. A larger size can make protein folding more complicated which might include non-functional fold states that the protein is prone to adapt without the guidance of chaperones. During proteotoxic stress, the quality control might be overwhelmed and cannot act on all proteins. Therefore, large proteins are prone for adapting non-functional fold states with exposed hydrophobic regions. Furthermore, common precipitants were shown to decrease the solubility of proteins with high molecular weight more than low weight proteins *in vitro* (Kramer *et al*, 2012). Since similar observation has been made in real cellular context and reduced *in vitro* conditions, it suggests that molecular weight *per se* makes protein

prone for aggregation. High molecular weight proteins are often composed of multiple individually folding domains; the domains are most often the results of gene duplications (Vogel *et al*, 2004). Unfolding and exposure of hydrophobic regions of only one domain could be sufficient enough to aggregate the whole protein.

The higher hydrophilicity (Figure 17A) and higher pI (Figure 17B) observed with aggregating proteins are probably related since high pI is caused by positively charged (and therefore hydrophilic) residues. An amino acid level analysis revealed that charged residues (except for negatively charged aspartate) are indeed more abundant in aggregating proteins (Figure 17C). However, the charged residues do not explain the more hydrophilic nature of aggregating proteins. The enrichment of more hydrophilic amino acids in aggregating proteins as well as depletion of hydrophobic ones is a clear trend among amino acids (Figure 17C).

When considering protein aggregation being driven by hydrophobic effect, the trend of increasing hydrophobic nature in aggregating proteins is interesting. One might expect the opposite outcome: more hydrophobic regions would “offer” more interaction surface between misfolded proteins to aggregate. As mention earlier, disordered regions *per se* are soluble (Uemura *et al*, 2018). Therefore, the enrichment of disordered regions in aggregators could related to the more hydrophilic nature observed in them.

In yeast, when comparing multiple stress conditions (that did not include heat shock), the general trend was that aggregating proteins tend to be more hydrophobic (Weids *et al*, 2016). This could mean that hydrophobicity is not a feature that would make proteins prone for aggregation, as with high molecular weight discussed above. However, it could also reflect the difference in aggregation between human and yeast cells. Interestingly, aggregation prone proteins in yeast differ also in their pI when compared to human cell upon heat shock. Here, aggregating proteins had higher pI while yeast cells had consistently lower pI in aggregates formed under different stress conditions

(Weids *et al*, 2016). Similarly, as with the hydrophobicity, this could mean that different proteins aggregate in different stress conditions or in different organisms.

## 7. Slow disaggregation in metazoans

Full disaggregation of heat shock-induced aggregates in yeast *in vivo* occurs approximately within an hour (Wallace *et al*, 2015); however, many proteins were disaggregated after 20 minutes. It should also be noted that the full disaggregation could have happen much faster since the time points to follow disaggregation in the experiment were 20 and 60 minutes. Similar time-scale for full disaggregation was observed for luciferase by human Hsp70-based disaggregase *in vitro* (Nillegoda *et al*, 2015); the activity of heat-aggregated luciferase was completely recovered after 60 minutes. However, the disaggregation *in situ* of human endogenous aggregates were still on their way of being fully disaggregated after five hours of recovery (Figure 20B). Although, few fastest disaggregating proteins do return to the control levels of solubility, i.e. are fully disaggregated (Figure 23). Extrapolating the disaggregation in Figure 20B suggests that the full disaggregation would take approximately 12 hours (assuming that disaggregation would follow the linear trend and full disaggregation would be achieved). These results shows that, although fast in the reduced *in vitro* settings, in the biological context disaggregation in humans is much slower than in yeast.

The slow disaggregation observed here might not be specific just for human cells. Similar with what was found for human cells, disaggregation of luciferase in *in vitro* with nematode disaggregase was also fast (Kirstein *et al*, 2017, Nillegoda *et al*, 2015); the disaggregation reached plateau also approximately after 60 minutes, although, the recovery of heat-aggregated luciferase was not as complete (60-80%) as with human disaggregase. However, in *in vivo* assay, traces of heat shock-induced luciferase

aggregates were found even days after recovery from heat shock (Kirstein *et al*, 2017). This suggests that also nematode can have disaggregase that has the ability to re-solubilize aggregated proteins within a time scale of one hour; on cellular context the capacity is, however, much more limited.

During recovery from non-lethal heat shock, yeast cells are able to re-solubilize aggregates much faster than human cells or nematode. It should be noted that the nematode data is based on following the aggregation and disaggregation of expressed model protein. Yeast harbors a powerful Hsp100-family disaggregase system that can be more efficient than human Hsp70-based system (Rampelt *et al*, 2012). This could explain the different kinetics between human and yeast disaggregation. Hsp100 disaggregation system is also absent from nematode as well as from other metazoans (Nillegoda & Bukau, 2015). Therefore, it would be tempting to conclude that the absence of Hsp100-disaggregase has led to slower disaggregation in metazoan.

The very different disaggregation could also be related to the proteins that are re-solubilized. As discussed previously, proteins with quite different qualities might be prone for aggregation at least in yeast and human. How much the disaggregation differs if the aggregating proteins are more hydrophilic (human) or hydrophobic (yeast)? The composition of Hsp40s in human disaggregase has selectivity over the size of the aggregates (Nillegoda *et al*, 2015), suggesting that even the aggregate shape of one protein species can impact the disaggregation efficiency. Therefore, the physicochemical character of the protein being disaggregated could contribute to the disaggregation kinetics. If different species have different kinds of proteins aggregating, then the difference between diagggregation could be a direct consequence of that.

## 8. Soluble remnants of aggregating proteins are protected from aggregation

A strong stabilization was observed for sub-pool of proteins that remain soluble while another pool aggregates (Figure 32). The strong stabilization means that higher temperatures were needed to aggregate these soluble sub-pools. An interesting question is that did these sub-pools have the higher thermal stability already before the heat shock or was it induced by the heat shock? If heat shock would induce a stabilization of sub-pool of aggregating proteins, it could reflect a heat-induced mechanism to protect them from aggregation. It is important to note that the 2D-TPP assay is performed immediately after heat or mock shock. This eliminates the effects stemming from, for example, protein synthesis. Therefore, the stabilization would be an immediate response to the heat shock.

Thermal stabilization can reflect post-translational modification added to (or removed from) a protein (Huang *et al*, 2019, Potel *et al*, 2020, Smith *et al*, 2020). This might be reflected on the stabilization of MAPK kinases of stress-activated MAPK pathway (Figure 33) which are phosphorylated upon stress (Hotamisligil & Davis, 2016). Intra- or intermolecular interactions can also stabilize a protein (Becher *et al*, 2018, Pucci & Rومان, 2017). These effects might be linked to the destabilization of protein complexes related to protein quality control (Figure 35). It should be highlighted that the major limitation of 2D-TPP is that stability changes do not provide direct evidence for the events that led to (de)stabilization.

In the context of proteotoxic heat stress, it is very interesting that the soluble sub-pools of practically all aggregating proteins were systematically stabilized. What could be behind this wide-spread stabilization? One might expect that if a certain protein species would be prone for aggregation, it would aggregate more or less completely. As

discussed earlier, some proteins aggregate to different extent than others (which was also negatively correlating to disaggregation rates). This indicates that complete aggregation did not take place at least in the used conditions. The wide-spread stabilization could be the result of kinases or other enzymes making post-translational modifications. In addition, it would be tempting to speculate the actions of a chaperone network, such as the epichaperome (Rodina *et al*, 2016), that would bind and stabilize a sub-pool of unstable proteins. If the aggregating proteins unfold and expose their hydrophobic regions, this might recruit chaperones. However, as the chaperone pool probably is not capable of handling such a wide-spread unfolding, only a sub-pool of the unfolded proteins could be “saved” by the chaperones. As chaperones are quite unspecific, they would bind and stabilize a wide variety of different proteins with exposed hydrophobic regions. This could explain why approximately all aggregating proteins were stabilized. At the same time, proteins might not be completely “saved” by the chaperones because there is a limited supply of chaperones. Of course, if a protein would be completely prevented from aggregation by chaperone binding, no loss in solubility would not be seen. Although, this might lead to higher thermal stability (which could be detected but would be challenging to differentiate from other proteins that might be stabilized by other mechanisms, such as phosphorylation). As discussed earlier, aggregation in nucleus can relate to chromatin binding (or aggregation on chromatin) that could lead to protein stabilization. The available space on chromatin could limit the number of proteins able to “escape” aggregation and results in protection of only a small sub-pool of each protein species. The protection from aggregation in this context might be related to functional DNA-binding (e.g. chromatin remodelling) rather than trying to save the protein from aggregation.

## REFERENCES

- Aebersold R, Goodlett DR (2001) Mass spectrometry in proteomics. *Chemical reviews* 101: 269-95
- Aggarwal S, Talukdar NC, Yadav AK (2019) Advances in Higher Order Multiplexing Techniques in Proteomics. *J Proteome Res* 18: 2360-2369
- Alberti S, Mateju D, Mediani L, Carra S (2017) Granulostasis: Protein Quality Control of RNP Granules. *Front Mol Neurosci* 10: 84
- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181: 223-30
- Aprile-Garcia F, Tomar P, Hummel B, Khavaran A, Sawarkar R (2019) Nascent-protein ubiquitination is required for heat shock-induced gene downregulation in human cells. *Nat Struct Mol Biol* 26: 137-146
- Avellaneda MJ, Franke KB, Sunderlikova V, Bukau B, Mogk A, Tans SJ (2020) Processive extrusion of polypeptide loops by a Hsp100 disaggregase. *Nature* 578: 317-320
- Balchin D, Hayer-Hartl M, Hartl FU (2016) In vivo aspects of protein folding and quality control. *Science* 353: aac4354
- Bantscheff M, Lemeer S, Savitski MM, Kuster B (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* 404: 939-65
- Bauerlein FJB, Saha I, Mishra A, Kalemanov M, Martinez-Sanchez A, Klein R, Dudanova I, Hipp MS, Hartl FU, Baumeister W, Fernandez-Busnadiego R (2017) In Situ Architecture and Cellular Interactions of PolyQ Inclusions. *Cell* 171: 179-187.e10
- Becher I, Andres-Pons A, Romanov N, Stein F, Schramm M, Baudin F, Helm D, Kurzawa N, Mateus A, Mackmull MT, Typas A, Muller CW, Bork P, Beck M, Savitski MM (2018) Pervasive Protein Thermal Stability Variation during the Cell Cycle. *Cell* 173: 1495-1507 e18
- Becher I, Werner T, Doce C, Zaal EA, Togel I, Khan CA, Rueger A, Muelbaier M, Salzer E, Berkers CR, Fitzpatrick PF, Bantscheff M, Savitski MM (2016) Thermal profiling reveals phenylalanine hydroxylase as an off-target of panobinostat. *Nat Chem Biol* 12: 908-910
- Biamonti G, Vourc'h C (2010) Nuclear stress bodies. *Cold Spring Harb Perspect Biol* 2: a000695

- Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M (2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 28: 420-435
- Boisvert FM, Lamond AI (2010) p53-Dependent subcellular proteome localization following DNA damage. *Proteomics* 10: 4087-97
- Bolognesi B, Lorenzo Gotor N, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, Lehner B (2016) A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell reports* 16: 222-231
- Braakman I, Hebert DN (2013) Protein folding in the endoplasmic reticulum. *Cold Spring Harb Perspect Biol* 5: a013201
- Bracher A, Verghese J (2015) The nucleotide exchange factors of Hsp70 molecular chaperones. *Front Mol Biosci* 2: 10
- Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Julicher F, Hyman AA (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324: 1729-32
- Buchan JR, Parker R (2009) Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* 36: 932-41
- Buchberger A, Bukau B, Sommer T (2010) Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol Cell* 40: 238-52
- Chahrour O, Cobice D, Malone J (2015) Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *Journal of pharmaceutical and biomedical analysis* 113: 2-20
- Chiti F, Dobson CM (2017) Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* 86: 27-68
- Collier NC, Heuser J, Levy MA, Schlesinger MJ (1988) Ultrastructural and biochemical analysis of the stress granule in chicken embryo fibroblasts. *J Cell Biol* 106: 1131-9
- Collier NC, Schlesinger MJ (1986) The dynamic state of heat shock proteins in chicken embryo fibroblasts. *J Cell Biol* 103: 1495-507
- Consortium U (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic acids research* 47: D506-d515
- Corboy MJ, Thomas PJ, Wigley WC (2005) Aggresome formation. *Methods Mol Biol* 301: 305-27
- Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem* 65: 801-47
- Cox J, Mann M (2007) Is proteomics the new genomics? *Cell* 130: 395-8



David DC, Ollikainen N, Trinidad JC, Cary MP, Burlingame AL, Kenyon C (2010) Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS biology* 8: e1000450

Davis RJ (2000) Signal transduction by the JNK group of MAP kinases. *Cell* 103: 239-52

De Maio A, Santoro MG, Tanguay RM, Hightower LE (2012) Ferruccio Ritossa's scientific legacy 50 years after his discovery of the heat shock response: a new view of biology, a new society, and a new journal. *Cell Stress Chaperones* 17: 139-43

Deane CA, Brown IR (2017) Components of a mammalian protein disaggregation/refolding machine are targeted to nuclear speckles following thermal stress in differentiated human neuronal cells. *Cell Stress Chaperones* 22: 191-200

Dephoure N, Gygi SP (2012) Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Science signaling* 5: rs2

Dikic I (2017) Proteasomal and Autophagic Degradation Systems. *Annu Rev Biochem* 86: 193-224

Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 25: 331-9

Doherty MK, Hammond DE, Clague MJ, Gaskell SJ, Beynon RJ (2009) Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J Proteome Res* 8: 104-12

Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312: 212-7

Doyle SM, Genest O, Wickner S (2013) Protein rescue from aggregates by powerful molecular chaperone machines. *Nat Rev Mol Cell Biol* 14: 617-29

Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134: 341-52

Drummond IA, McClure SA, Poenie M, Tsien RY, Steinhardt RA (1986) Large changes in intracellular pH and calcium observed during heat shock are not responsible for the induction of heat shock proteins in *Drosophila melanogaster*. *Mol Cell Biol* 6: 1767-75

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6: 197-208

Eliuk S, Makarov A (2015) Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual review of analytical chemistry (Palo Alto, Calif)* 8: 61-80

Espay AJ, Vizcarra JA, Marsili L, Lang AE, Simon DK, Merola A, Josephs KA, Fasano A, Morgante F, Savica R, Greenamyre JT, Cambi F, Yamasaki TR, Tanner CM, Gan-Or

- Z, Litvan I, Mata IF, Zabetian CP, Brundin P, Fernandez HH et al. (2019) Revisiting protein aggregation as pathogenic in sporadic Parkinson and Alzheimer diseases. *Neurology* 92: 329-337
- Fierro-Monti I, Racle J, Hernandez C, Waridel P, Hatzimanikatis V, Quadroni M (2013) A novel pulse-chase SILAC strategy measures changes in protein decay and synthesis rates induced by perturbation of proteostasis with an Hsp90 inhibitor. *PLoS One* 8: e80423
- Franken H, Mathieson T, Childs D, Sweetman GM, Werner T, Togel I, Doce C, Gade S, Bantscheff M, Drewes G, Reinhard FB, Huber W, Savitski MM (2015) Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat Protoc* 10: 1567-93
- Frottin F, Schueder F, Tiwary S, Gupta R, Korner R, Schlichthaerle T, Cox J, Jungmann R, Hartl FU, Hipp MS (2019) The nucleolus functions as a phase-separated protein quality control compartment. *Science*
- Funk RHW, Nagel F, Wonka F, Krinke HE, Gölfert F, Hofer A (1999) Effects of heat shock on the functional morphology of cell organelles observed by video-enhanced microscopy. *The Anatomical Record* 255: 458-464
- Fuxreiter M, Simon I, Bondos S (2011) Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem Sci* 36: 415-23
- Galluzzi L, Baehrecke EH, Ballabio A, Boya P, Bravo-San Pedro JM, Cecconi F, Choi AM, Chu CT, Codogno P, Colombo MI, Cuervo AM, Debnath J, Deretic V, Dikic I, Eskelinen EL, Fimia GM, Fulda S, Gewirtz DA, Green DR, Hansen M et al. (2017) Molecular definitions of autophagy and related processes. *Embo j*
- Gatto L, Lilley KS (2012) MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* 28: 288-9
- Gibson JS, Ellory JC (2002) Membrane transport in sickle cell disease. *Blood cells, molecules & diseases* 28: 303-14
- Gillet LC, Leitner A, Aebersold R (2016) Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annual review of analytical chemistry (Palo Alto, Calif)* 9: 449-72
- Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiological reviews* 82: 373-428
- Griffiths WJ, Wang Y (2009) Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chemical Society reviews* 38: 1882-96
- Grousl T, Ivanov P, Malcova I, Pompach P, Frydlova I, Slaba R, Senohrabkova L, Novakova L, Hasek J (2013) Heat shock-induced accumulation of translation elongation

and termination factors precedes assembly of stress granules in *S. cerevisiae*. PLoS One 8: e57083

Hartl FU (2016) Cellular Homeostasis and Aging. Annu Rev Biochem 85: 1-4

Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475: 324-32

Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol 16: 574-81

Hergeth SP, Schneider R (2015) The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep 16: 1439-53

Hill SM, Hanzen S, Nystrom T (2017) Restricted access: spatial sequestration of damaged proteins during stress and aging. EMBO Rep 18: 377-391

Hinkson IV, Elias JE (2011) The dynamic state of protein turnover: It's about time. Trends Cell Biol 21: 293-303

Hjerpe R, Bett JS, Keuss MJ, Solovyova A, McWilliams TG, Johnson C, Sahu I, Varghese J, Wood N, Wightman M, Osborne G, Bates GP, Glickman MH, Trost M, Knebel A, Marchesi F, Kurz T (2016) UBQLN2 Mediates Autophagy-Independent Protein Aggregate Clearance by the Proteasome. Cell 166: 935-49

Hosp F, Gutierrez-Angel S, Schaefer MH, Cox J, Meissner F, Hipp MS, Hartl FU, Klein R, Dudanova I, Mann M (2017) Spatiotemporal Proteomic Profiling of Huntington's Disease Inclusions Reveals Widespread Loss of Protein Function. Cell reports 21: 2291-2303

Hotamisligil GS, Davis RJ (2016) Cell Signaling and Stress Responses. Cold Spring Harb Perspect Biol 8

Huang JX, Lee G, Cavanaugh KE, Chang JW, Gardel ML, Moellering RE (2019) High throughput discovery of functional protein modifications by Hotspot Thermal Profiling. Nat Methods 16: 894-901

Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 Suppl 1: S96-104

Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. Mol Syst Biol 10: 757

Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C (2015) The GOA database: gene Ontology annotation updates for 2015. Nucleic acids research 43: D1057-63

- Ibstedt S, Sideri TC, Grant CM, Tamas MJ (2014) Global analysis of protein aggregation in yeast during physiological conditions and arsenite stress. *Biology open* 3: 913-23
- Ivanov P, Kedersha N, Anderson P (2018) Stress Granules and Processing Bodies in Translational Control. *Cold Spring Harb Perspect Biol*
- Johansen T, Lamark T (2014) Selective autophagy mediated by autophagic adapter proteins. *Autophagy* 7: 279-296
- Johnston JA, Ward CL, Kopito RR (1998) Aggresomes: a cellular response to misfolded proteins. *J Cell Biol* 143: 1883-98
- Jolly C, Lakhotia SC (2006) Human sat III and Drosophila hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells. *Nucleic acids research* 34: 5508-14
- Jones RD, Gardner RG (2016) Protein quality control in the nucleus. *Current opinion in cell biology* 40: 81-89
- Kampinga HH, Craig EA (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* 11: 579-92
- Kanshin E, Kubiniok P, Thattikota Y, D'Amours D, Thibault P (2015) Phosphoproteome dynamics of *Saccharomyces cerevisiae* under heat shock and cold stress. *Mol Syst Biol* 11: 813
- Kirstein J, Arnsburg K, Scior A, Szlachcic A, Guilbride DL, Morimoto RI, Bukau B, Nillegoda NB (2017) In vivo properties of the disaggregase function of J-proteins and Hsc70 in *Caenorhabditis elegans* stress and aging. *Aging cell* 16: 1414-1424
- Knowles TP, Vendruscolo M, Dobson CM (2014) The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* 15: 384-96
- Kornberg RD (1999) Eukaryotic transcriptional control. *Trends Cell Biol* 9: M46-9
- Kovács D, Sigmond T, Hotzi B, Bohár B, Fazekas D, Deák V, Vellai T, Barna J (2019) HSF1Base: A Comprehensive Database of HSF1 (Heat Shock Factor 1) Target Genes. *International journal of molecular sciences* 20
- Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM (2012) Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophysical journal* 102: 1907-15
- Kuckelkorn U, Knuehl C, Boes-Fabian B, Drung I, Kloetzel PM (2000) The effect of heat shock on 20S/26S proteasomes. *Biological chemistry* 381: 1017-23
- Kumar R, Neuser N, Tyedmers J (2017) Hitchhiking vesicular transport routes to the vacuole: Amyloid recruitment to the Insoluble Protein Deposit (IPOD). *Prion* 11: 71-81

Kundra R, Dobson CM, Vendruscolo M (2020) A Cell- and Tissue-Specific Weakness of the Protein Homeostasis System Underlies Brain Vulnerability to Protein Aggregation. *iScience* 23: 100934

Leuenberger P, Gansch S, Kahraman A, Cappelletti V, Boersema PJ, von Mering C, Claassen M, Picotti P (2017) Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* 355

Li J, Van Vranken JG, Pontano Vaite L, Schweppe DK, Huttlin EL, Etienne C, Nandhikonda P, Viner R, Robitaille AM, Thompson AH, Kuhn K, Pike I, Bomgarden RD, Rogers JC, Gygi SP, Paulo JA (2020) TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat Methods* 17: 399-404

Lopez T, Dalton K, Frydman J (2015) The Mechanism and Function of Group II Chaperonins. *J Mol Biol* 427: 2919-30

Mahat Dig B, Salamanca HH, Duarte Fabiana M, Danko Charles G, Lis John T (2016) Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular Cell* 62: 63-78

Mateus A, Bobonis J, Kurzawa N, Stein F, Helm D, Hevler J, Typas A, Savitski MM (2018) Thermal proteome profiling in bacteria: probing protein state in vivo. *Mol Syst Biol* 14: e8242

Mateus A, Maatta TA, Savitski MM (2017) Thermal proteome profiling: unbiased assessment of protein state through heat-induced stability changes. *Proteome science* 15: 13

Mathieson T, Franken H, Kosinski J, Kurzawa N, Zinn N, Sweetman G, Poeckel D, Ratnu VS, Schramm M, Becher I, Steidel M, Noh KM, Bergamini G, Beck M, Bantscheff M, Savitski MM (2018) Systematic analysis of protein turnover in primary cells. *Nat Commun* 9: 689

Mayer MP (2010) Gymnastics of molecular chaperones. *Mol Cell* 39: 321-31

Mayer MP, Gierasch LM (2019) Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *J Biol Chem* 294: 2085-2097

Miller SB, Mogk A, Bukau B (2015) Spatially organized aggregation of misfolded proteins as cellular stress defense strategy. *J Mol Biol* 427: 1564-74

Moggridge S, Sorensen PH, Morin GB, Hughes CS (2018) Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. *J Proteome Res* 17: 1730-1740

Mogk A, Bukau B, Kampinga HH (2018) Cellular Handling of Protein Aggregates by Disaggregation Machines. *Mol Cell* 69: 214-226

- Mogk A, Kummer E, Bukau B (2015) Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation. *Front Mol Biosci* 2: 22
- Mogk A, Ruger-Herreros C, Bukau B (2019) Cellular Functions and Mechanisms of Action of Small Heat Shock Proteins. *Annual review of microbiology*
- Muhlhofer M, Berchtold E, Stratil CG, Csaba G, Kunold E, Bach NC, Sieber SA, Haslbeck M, Zimmer R, Buchner J (2019) The Heat Shock Response in Yeast Maintains Protein Homeostasis by Chaperoning and Replenishing Proteins. *Cell reports* 29: 4593-4607.e8
- Nillegoda NB, Bukau B (2015) Metazoan Hsp70-based protein disaggregases: emergence and mechanisms. *Front Mol Biosci* 2: 57
- Nillegoda NB, Kirstein J, Szlachcic A, Berynskyy M, Stank A, Stengel F, Arnsburg K, Gao X, Scior A, Aebersold R, Guilbride DL, Wade RC, Morimoto RI, Mayer MP, Bukau B (2015) Crucial HSP70 co-chaperone complex unlocks metazoan protein disaggregation. *Nature* 524: 247-51
- Nillegoda NB, Stank A, Malinverni D, Alberts N, Szlachcic A, Barducci A, De Los Rios P, Wade RC, Bukau B (2017) Evolution of an intricate J-protein network driving protein disaggregation in eukaryotes. *Elife* 6
- Nillegoda NB, Wentink AS, Bukau B (2018) Protein Disaggregation in Multicellular Organisms. *Trends Biochem Sci* 43: 285-300
- Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ (2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* 57: 936-47
- O'Connell JD, Tsechansky M, Royall A, Boutz DR, Ellington AD, Marcotte EM (2014) A proteomic survey of widespread protein aggregation in yeast. *Mol Biosyst* 10: 851-61
- Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, Gough J (2013) D(2)P(2): database of disordered protein predictions. *Nucleic acids research* 41: D508-16
- Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1: 376-86
- Ori A, Iskar M, Buczak K, Kastiris P, Parca L, Andres-Pons A, Singer S, Bork P, Beck M (2016) Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome biology* 17: 47
- Osorio DR-V, P.; Torres, R. (2015) Peptides: A Package for Data Mining of Antimicrobial Peptides. *The R Journal* 7: 4-14

Parsell DA, Kowal AS, Singer MA, Lindquist S (1994) Protein disaggregation mediated by heat-shock protein Hsp104. *Nature* 372: 475-8

Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoyanov S, Mahamid J, Saha S, Franzmann TM, Pozniakovski A, Poser I, Maghelli N, Royer LA, Weigert M, Myers EW, Grill S, Drechsel D, Hyman AA, Alberti S (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162: 1066-77

Pelham HR (1984) Hsp70 accelerates the recovery of nucleolar morphology after heat shock. *Embo j* 3: 3095-100

Potel CM, Kurzawa N, Becher I, Typas A, Mateus A, Savitski MM (2020) Impact of phosphorylation on thermal stability of proteins. *bioRxiv*: 2020.01.14.903849

Pucci F, Rooman M (2017) Physical and molecular bases of protein thermal stability and cold adaptation. *Current opinion in structural biology* 42: 117-128

Rampelt H, Kirstein-Miles J, Nillegoda NB, Chi K, Scholz SR, Morimoto RI, Bukau B (2012) Metazoan Hsp70 machines use Hsp110 to power protein disaggregation. *EMBO J* 31: 4221-35

Reinhard FB, Eberhard D, Werner T, Franken H, Childs D, Doce C, Savitski MF, Huber W, Bantscheff M, Savitski MM, Drewes G (2015) Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat Methods* 12: 1129-31

Riback JA, Katanski CD, Kear-Scott JL, Pilipenko EV, Rojek AE, Sosnick TR, Drummond DA (2017) Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* 168: 1028-1040 e19

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43: e47

Ritossa F (1962) A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia* 18: 571-573

Ritossa F (1996) Discovery of the heat shock response. *Cell Stress Chaperones* 1: 97-8

Rodina A, Wang T, Yan P, Gomes ED, Dunphy MP, Pillarsetty N, Koren J, Gerecitano JF, Taldone T, Zong H, Caldas-Lopes E, Alpaugh M, Corben A, Riolo M, Beattie B, Pressl C, Peter RI, Xu C, Trondl R, Patel HJ et al. (2016) The epichaperome is an integrated chaperome network that facilitates tumour survival. *Nature* 538: 397-401

Roque A, Iloro I, Ponte I, Arrondo JL, Suau P (2005) DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. *J Biol Chem* 280: 32141-7

Saad S, Cereghetti G, Feng Y, Picotti P, Peter M, Dechant R (2017) Reversible protein aggregation is a protective mechanism to ensure cell cycle restart after stress. *Nature cell biology* 19: 1202-1213

- Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* 14: 630-42
- Sanchez Y, Lindquist SL (1990) HSP104 required for induced thermotolerance. *Science* 248: 1112-5
- Sang P, Yang Q, Du X, Yang N, Yang LQ, Ji XL, Fu YX, Meng ZH, Liu SQ (2016) Effect of the Solvent Temperatures on Dynamics of Serine Protease Proteinase K. *International journal of molecular sciences* 17: 254
- Savitski MM, Reinhard FB, Franken H, Werner T, Savitski MF, Eberhard D, Martinez Molina D, Jafari R, Dovega RB, Klaeger S, Kuster B, Nordlund P, Bantscheff M, Drewes G (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* 346: 1255784
- Savitski MM, Zinn N, Faeltz-Savitski M, Poeckel D, Gade S, Becher I, Muelbaier M, Wagner AJ, Strohmmer K, Werner T, Melchert S, Petretich M, Rutkowska A, Vappiani J, Franken H, Steidel M, Sweetman GM, Gilan O, Lam EYN, Dawson MA et al. (2018) Multiplexed Proteome Dynamics Profiling Reveals Mechanisms Controlling Protein Homeostasis. *Cell*
- Schamhart DHJ, Walraven HSv, Wiegant FAC, Linnemans WAM, Rijn Jv, Berg Jvd, Wijk Rv (1984) Thermotolerance in Cultured Hepatoma Cells: Cell Viability, Cell Morphology, Protein Synthesis, and Heat-Shock Proteins. *Radiation Research* 98: 82-95
- Schramm FD, Schroeder K, Jonas K (2019) Protein aggregation in bacteria. *FEMS Microbiology Reviews* 44: 54-72
- Schwanhauser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* 473: 337-42
- Scigelova M, Hornshaw M, Giannakopoulos A, Makarov A (2011) Fourier transform mass spectrometry. *Mol Cell Proteomics* 10: M111.009431
- Scior A, Juenemann K, Kirstein J (2016) Cellular strategies to cope with protein aggregation. *Essays Biochem* 60: 153-161
- Sims RJ, 3rd, Mandal SS, Reinberg D (2004) Recent highlights of RNA-polymerase-II-mediated transcription. *Current opinion in cell biology* 16: 263-71
- Smith IR, Hess KN, Bakhtina AA, Valente AS, Rodríguez-Mias RA, Villén J (2020) Identification of phosphosites that alter protein thermal stability. *bioRxiv*: 2020.01.14.904300
- Specht S, Miller SB, Mogk A, Bukau B (2011) Hsp42 is required for sequestration of protein aggregates into deposition sites in *Saccharomyces cerevisiae*. *J Cell Biol* 195: 617-29



- Sridharan S, Kurzawa N, Werner T, Gunthner I, Helm D, Huber W, Bantscheff M, Savitski MM (2019) Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat Commun* 10: 1155
- Srinivas UK, Revathi CJ (1993) Changes in cytoplasmic pH upon heat shock in embryonic and adult rat liver cells. *Journal of Biosciences* 18: 175-186
- Sui X, Pires DEV, Ormsby AR, Cox D, Nie S, Vecchi G, Vendruscolo M, Ascher DB, Reid GE, Hatters DM (2020) Widespread remodeling of proteome solubility in response to different protein homeostasis stresses. *Proc Natl Acad Sci U S A*
- Tan CSH, Go KD, Bisteau X, Dai L, Yong CH, Prabhu N, Ozturk MB, Lim YT, Sreekumar L, Lenggqvist J, Tergaonkar V, Kaldis P, Sobota RM, Nordlund P (2018) Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. *Science* 359: 1170-1177
- Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75: 1895-904
- Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Bjork L, Breckels LM, Backstrom A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S et al. (2017) A subcellular map of the human proteome. *Science* 356
- Tyedmers J, Mogk A, Bukau B (2010) Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* 11: 777-88
- Uemura E, Niwa T, Minami S, Takemoto K, Fukuchi S, Machida K, Imataka H, Ueda T, Ota M, Taguchi H (2018) Large-scale aggregation analysis of eukaryotic proteins reveals an involvement of intrinsically disordered regions in protein folding. *Sci Rep* 8: 678
- Vabulas RM, Raychaudhuri S, Hayer-Hartl M, Hartl FU (2010) Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harb Perspect Biol* 2: a004390
- Valastyan JS, Lindquist S (2014) Mechanisms of protein-folding diseases at a glance. *Dis Model Mech* 7: 9-14
- Vaquier-Alicea J, Diamond MI (2019) Propagation of Protein Aggregation in Neurodegenerative Diseases. *Annual Review of Biochemistry* 88: 785-810
- Vecchi G, Sormanni P, Mannini B, Vandelli A, Tartaglia GG, Dobson CM, Hartl FU, Vendruscolo M (2019) Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proc Natl Acad Sci U S A*
- Vihervaara A, Mahat DB, Guertin MJ, Chu T, Danko CG, Lis JT, Sistonen L (2017) Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nat Commun* 8: 255

- Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. *Current opinion in structural biology* 14: 208-16
- Voos W, Jaworek W, Wilkening A, Bruderek M (2016) Protein quality control at the mitochondrion. *Essays Biochem* 60: 213-225
- Vuzman D, Levy Y (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosyst* 8: 47-57
- Wallace EW, Kear-Scott JL, Pilipenko EV, Schwartz MH, Laskowski PR, Rojek AE, Katanski CD, Riback JA, Dion MF, Franks AM, Airoidi EM, Pan T, Budnik BA, Drummond DA (2015) Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* 162: 1286-98
- Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, Ciryam P, Morimoto RI, Dobson CM, Vendruscolo M, Mann M, Hartl FU (2015) Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*. *Cell* 161: 919-32
- Wang L, Schubert D, Sawaya MR, Eisenberg D, Riek R (2010) Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew Chem Int Ed Engl* 49: 3904-8
- Weber SC, Brangwynne CP (2015) Inverse size scaling of the nucleolus by a concentration-dependent phase transition. *Curr Biol* 25: 641-6
- Weids AJ, Ibstedt S, Tamas MJ, Grant CM (2016) Distinct stress conditions result in aggregation of proteins with similar properties. *Sci Rep* 6: 24554
- Welch WJ, Feramisco JR (1984) Nuclear and nucleolar localization of the 72,000-dalton heat shock protein in heat-shocked mammalian cells. *J Biol Chem* 259: 4501-13
- Welch WJ, Suhan JP (1985) Morphological study of the mammalian stress response: characterization of changes in cytoplasmic organelles, cytoskeleton, and nucleoli, and appearance of intranuclear actin filaments in rat fibroblasts after heat-shock treatment. *Journal of Cell Biology* 101: 1198-1211
- Werner T, Sweetman G, Savitski MF, Mathieson T, Bantscheff M, Savitski MM (2014) Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal Chem* 86: 3594-601
- Wolf DH, Hilt W (2004) The proteasome: a proteolytic nanomachine of cell regulation and waste disposal. *Biochim Biophys Acta* 1695: 19-31
- Wright ES (2016) Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal* 8: 352-359
- Yamamoto K, Furukawa MT, Fukumura K, Kawamura A, Yamada T, Suzuki H, Hirose T, Sakamoto H, Inoue K (2016) Control of the heat stress-induced alternative splicing of a subset of genes by hnRNP K. *Genes Cells* 21: 1006-14

Zanton SJ, Pugh BF (2006) Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes Dev* 20: 2250-65

Zecha J, Meng C, Zolg DP, Samaras P, Wilhelm M, Kuster B (2018) Peptide level turnover measurements enable the study of proteoform dynamics. *Mol Cell Proteomics*

Zwirowski S, Klosowska A, Obuchowski I, Nillegoda NB, Pirog A, Zietkiewicz S, Bukau B, Mogk A, Liberek K (2017) Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. *EMBO J* 36: 783-796